BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/10, 15/13, C12Q 1/68
// C12N 15/62

A1

(11) International Publication Number: WO 92/15678

(43) International Publication Date: 17 September 1992 (17.09.92)

(21) International Application Number: PCT/US92/01475

(22) International Filing Date: 27 February 1992 (27.02.92)

(30) Priority data:
663,442 1 March 1991 (01.03.91) US

(71) Applicant: STRATAGENE [US/US]; 11099 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventor: SORGE, Joseph, A. ; P.O. Box 9437, Rancho Santa Fe, CA 92067 (US).

(74) Agents: SIMPSON, Andrew, H. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, Suite 1600, Newport Beach, CA 92660 (US).

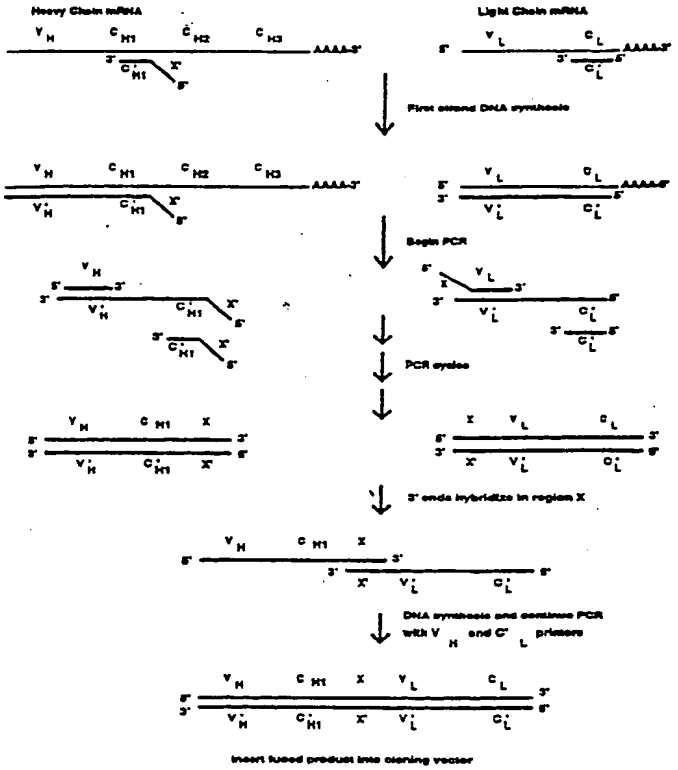
(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: PCR GENERATED DICISTRONIC DNA MOLECULES FOR PRODUCING ANTIBODIES

(57) Abstract

A method of producing dicistronic DNA molecules each having upstream and downstream cistrons respectively coding for the first and second polypeptides of a heterodimeric receptor. Kits including, in separate containers, the primers and/or vectors of the invention in amounts sufficient to produce and/or express the dicistronic DNA molecules.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

PCR GENERATED DICISTRONIC DNA MOLECULES FOR PRODUCING ANTIBODIES

Technical Field

The present invention relates to a method for producing a library of dicistronic DNA molecules useful in expressing heterodimeric receptors, such as antibodies, T cell receptors and the like.

Background

The expression of antibody libraries in bacteria has opened up new ways to uncover monoclonal antibody specificities. The antigen binding domain of the antibody is composed of a heavy and a light chain. These chains are each encoded by separate genes. To reconstruct a complete binding domain in bacteria, both heavy and light chain coding sequences are typically coexpressed, which involves two cloning steps, one for the heavy chain and one for the light. This is generally accomplished by either inserting both heavy and light chain coding sequences into one vector, or by first making separate heavy and light chain libraries and recombining the genomes to make a combinatorial library encoding random combinations of the heavy and light sequences. In either case, the need to clone two separate DNA fragments is cumbersome and, therefore, a method that could fuse both heavy and light chain sequences together prior to vector ligation would be desirable.

Brief Description of the Invention

The present invention contemplates a method of producing dicistronic DNA molecules each having upstream and downstream cistrons respectively coding for first and second polypeptides of a heterodimeric protein, such as a receptor. The method comprises the

following steps:

(A) Forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, first polypeptide-encoding genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer. The inside first gene primer has a 3'-terminal priming portion and, preferably, a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a first gene.

(B) Subjecting the first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form.

(C) Forming a second PCR admixture by combining, in a PCR buffer, second polypeptide-encoding genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer. The inside gene primer has a 3'-terminal priming portion and, preferably, a 5'-terminal hybridizing portion complementary to a hybridizable portion of the 5'-terminal non-priming portion of the first inside gene primer. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a second polypeptide-coding gene.

The first and second inside primers, when hybridized, form a duplex that codes for a double-stranded cistronic bridge that links the upstream and downstream cistrons. One strand of the bridge codes for (i) at least one stop codon in the same reading frame as said upstream cistron, (ii) signals for the initiation of translation of the downstream in cistron. Preferably, such signals include a ribosome binding site downstream from the stop codon, and at

least one translation initiation codon in the same reading frame as the downstream cistron, the initiation codon being located downstream from the ribosome binding site.

(D) Subjecting the second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form.

(E) Separating the double stranded DNA homologs produced in steps (B) and (D).

(F) Hybridizing the separated strands of step (E) to form internally-primed duplexes.

(G) Subjecting the internally-primed duplexes to conditions for primer extension to produce a dicistronic DNA molecule. Each of the dicistronic DNA molecules produced contains a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by the cistronic bridge. The upstream cistron comprises one of the first polypeptide- or second polypeptide-coding DNA homologs. The downstream cistron comprises the other of the first polypeptide- or second polypeptide-coding DNA homologs.

Preferably, steps (A)-(D) are performed concurrently in one reaction vessel.

Preferably, the polypeptide-encoding genes of steps (A) and (B) are present in respective repertoires of conserved genes. When used, the repertoires of steps (A) and (C) are usually formed by isolating mRNA from at least about 10^3, preferably at least about 10^7 lymphocytes. It is preferred that the repertoire of first polypeptide genes comprises at least 10^5 different first polypeptide genes, and that the repertoire of second polypeptide genes comprises at least 10^5 different second polypeptide genes.

However, it should be noted that the method of the present invention can be used to operatively link for polycistronic expression any two genes. Thus, this invention can be used to physically link two genes from a single cell, such as a B cell, T cell, and the like, and thereby take advantage of a native immune system's ability to select operative gene pairs from the immunological repertoire. Similarly, operative gene pairs, i.e., a pair of genes encoding a heterodimeric receptor, from cells such as hybridomas, quadromas and the like, can be physically linked using the method of this invention.

Preferably the method further comprises step (H) wherein the dicistronic DNA molecules are PCR amplified by combining them with the outside first gene primer and the outside second gene primer to form a third PCR admixture. The third PCR admixture is then subjected to a plurality of PCR thermocycles. When a repertoire of first and/or second polypeptide-encoding genes is used, an amplified library of dicistronic DNA molecules is produced.

In preferred embodiments, the amplified products of step (H) are operatively linked for expression to a vector, preferably a phage vector. Preferably, the steps for operatively linking the dicistronic DNA molecules to a vector and isolating a recombinant vector that expresses a desired heterodimeric receptor include the following:

(i) Preparation of vector DNA and the dicistronic DNA molecules by cleavage with appropriate restriction enzyme(s) to form cohesive termini.

(ii) Ligation of the digested vector with the dicistronic DNA molecules via the cohesive termini.

(iii) Packaging of the ligated DNA (rDNA) into bacteriophage particles that can form plaques on appropriate bacterial hosts.

(iv) Identification of recombinant bacteriophages carrying the desired dicistronic DNA molecules.

(v) Plaque purification of selected recombinant bacteriophages.

Where the heterodimeric receptor is an antibody, the outside first gene primer hybridizes to a framework, leader or promoter region of a V_H immunoglobulin gene, and the outside second gene primer hybridizes to a J_L, constant or framework region, of a V_L immunoglobulin gene. The 3'-terminal priming portion of the inside first gene primer hybridizes to a J_H, hinge, constant, or framework region of a V_H immunoglobulin gene, and the 3'-terminal priming portion of the inside second gene primer hybridizes to a framework, leader or promoter region of a V_L immunoglobulin gene.

In another embodiment, a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, is produced by the following steps:

(A) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) V_H genes,

(ii) V_L genes,

(iii) an outside V_H gene primer

(iv) an outside V_L gene primer, and

(v) a linking primer having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer

extension product of one of the outside primers. The 5'-terminal primer template portion has a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

(B) Subjecting the PCR admixture of step (A) to a plurality of PCR thermocycles.

In preferred embodiments, the method further comprises steps (C)-(H) as follows:

(C) Subjecting the internally-primed duplexes to conditions for primer extension to produce dicistronic DNA molecules, each containing a V_H-coding sequence and a V_L-coding sequence linked by the cistronic bridge. The upstream cistron comprises one of the V_H- or V_L-coding DNA homologs, and the downstream cistron comprising the other of the V_H- V_L-coding DNA homologs.

(D) Operatively linking for expression the different dicistronic DNA molecules produced in step (C) to expression vectors, preferably phage vectors, thereby forming a plurality of V_{HL} expression vectors.

(E) Transforming a population of host cells, preferably E. coli compatible with the expression vector with a plurality of the V_{HL}-expression vectors to produce a transformed population of host cells whose members contain the V_{HL}-expression vectors.

(F) Culturing the transformed population under conditions for expressing the V_H and V_L polypeptides coded for by the dicistronic DNA molecules.

(G) Assaying the members of the transformed population for expression of an antibody molecule capable of binding a preselected ligand, thereby identifying transformants containing the dicistronic

DNA molecule.

(H) Segregating an identified transformant in step (G) from the population, thereby producing the isolated dicistronic DNA molecule.

Also contemplated are kits for producing a dicistronic DNA molecule as described herein. In one embodiment, the kit is an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, and a linking primer defining a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

Another contemplated kit comprises an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, an inside first polypeptide, preferably a V_H, gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a V_H gene. The kit also contains an inside second polypeptide, preferably a V_L, gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of the first polypeptide gene primer, the 3'-terminal priming portion of which comprises a

nucleotide sequence homologous to a conserved portion of a second polypeptide gene. The first polypeptide inside and second polypeptide inside primers, when hybridized, form a duplex that codes for a double-stranded DNA molecule containing the before described cistronic bridge for linking the upstream and downstream cistrons.

Brief Description of the Drawings

Figure 1 illustrates the principal structural features of an immunoglobulin molecule. The circled areas on the heavy and light chains represent the variable regions, (V_H) and (V_L), a heterodimeric polypeptide containing a biologically active (ligand binding) portion of that region, and genes coding for the individual polypeptides, are produced by the methods of the present invention.

Figure 2 contains three panels. Panel 2A illustrates various features of the heavy chain of human IgG (IgG1 subclass). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the presence of four domains, each containing an intrachain disulfide bond (S-S) and spanning approximately 110 amino acid residues. The symbol CHO stands for carbohydrate. The V region of the heavy (H) chain (V_H) resembles V_L in having three hypervariable complementarity determining regions (CDR'S) (not shown).

Panel 2B and 2C illustrate various features of a human kappa (K) chain. Numbering is from the N-terminus on the left to the C-terminus on the right. Note in Panel 2B the intrachain disulfide bond (S-S) spanning about the same number of amino acid residues in the V_L and C_L domains. Panel 2C shows the locations of the CDRs in the V_L domain. Segments

outside the CDR are the framework segments (FR).

Figure 3 illustrates a portion of the nucleotide base sequence of the 1661 base pair gene lamB sequence from residue number 250 to residue number 651. The base sequences are shown conventionally from left to right and in the direction of 5' terminus to 3' terminus using the single letter nucleotide base code (A = adenine, T = thymine, C = cytosine and G = guanine). The position of the nucleotide base sequence is indicated by the numbers in the left margin of the figure.

The reading frame of the structural lamB gene is indicated by placement of the deduced amino acid residue sequence of the lambda receptor protein for which it codes below the nucleotide sequence such that the triple letter code for each amino acid residue is located directly below the three bases (codon) coding for each residue. The residue sequence is shown conventionally from left to right and in the direction of amino terminus to carboxy terminus. The position of the amino acid residue sequence is indicated by the numbers in the right margin of the figure.

Figure 4 illustrates the strategy used to create immunoglobulin heavy and light chain PCR fusion products. RNA and DNA are represented by dotted and solid lines, respectively. Regions of the immunoglobulin heavy chain coding strand area designated V_H, C_H1, C_H2, and C_H3 correspond to those functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime (') following the symbol. Regions V_L and C_L are similarly labelled for the light chain. A region, X, unrelated to the natural immunoglobulin sequences is introduced into the fusion product by attaching X to the 5' ends of the C_H1' inside and V_L inside primers.

Figure 5 illustrates human fusion PCR inside primers. The heavy chain C_H1' inside primer sequence is written 3' to 5' and the light chain V_L inside primer sequence is written 5' to 3'. Note that it is not the primer strands that cross-prime to create the fusion molecule, but the complementary PCR product strands. Boxed nucleotides represent regions where the C_H1' primer hybridizes to the 3' end of C_H1 on human IgG heavy chain mRNA or where the V_L primer hybridizes to the 5' end of V_L framework-1 on human kappa light chain cDNA. Underlined sequences indicate the two stop codons. The italicized amino acid and nucleotides indicate changes in sequence from the original pelB leader sequence. The mouse fusion-PCR internal primers overlap in a similar manner.

Figure 6 illustrates the sequences of the synthetic DNAs inserted into Lambda ZAP to produce Lambda Zap II V_H (ImmunoZAP H) (Panel A) and Lambda Zap V_L (ImmunoZAP L) (Panel B) expression vectors. The various features required for these vectors to express the V_H and V_L-coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the V_H and V_L homologs to the expression vector. The V_H expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions heavy chain (V_H Backbone). This V_H Backbone is just upstream and in the proper reading as the V_H DNA homologs that are operatively linked into the Xho I and Spe I restriction sites. The V_L DNA homologs are operatively linked into the V_L sequence (Panel B) at the Sac I and Xba I restriction enzyme sites.

Figure 7 illustrates the major features of the bacterial expression vector Lambda Zap II V_H (ImmunoZAP H) (V_H- expression vector). The amino acids encoded by the synthetic DNA sequence from Figure 6A is shown at the top along with the T_3 polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The V_H DNA homologs were inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The V_H DNA homologs were inserted into the Xho I and Spe I restriction enzyme sites. The read through transcription produces the decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 8 illustrates, in Panels 8A and 8B, the major features of the bacterial expression vector Lambda ZAP II Modified V_H (Modified ImmunoZAP H) (V_H- expression vector) (IZ H). The amino acids encoded by the synthetic DNA sequence from Panel 8A is shown along with the T_3 polymerase promoter from Lambda ZAP II. The orientation of the insert in Lambda ZAP II is as presented. The insert was modified by the elimination of the Sac I site between the T_3 polymerase and Not I site and by the change of amino acids at the 5' end of the heavy chain from QVKL to QVQL (a lysine residue was changed to a glutamine residue). The V_H and V_L DNA homologs were inserted into the Xho I and Xba I cloning sites of the phagemid as described in Figure 7 and shown in Panel 8B. The modifications were made to create a fusion-PCR library from hybridoma RNA, to overcome decreased efficiency of secretion of positively charged amino acids in the amino terminus of the protein. Inouye et al., Proc. Natl. Acad. Sci., USA, 85;7685-7689 (1988), and to

make the V_L Sac I cloning site a unique restriction site.

Figure 9 illustrates the major features of the bacterial expression vector Lambda Zap II V_L (ImmunoZAP L) (V_L expression vector). The amino acids encoded by the synthetic DNA sequence shown in Figure 6B is shown at the top along with the T_3 polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The V_L DNA homologs are inserted into the Sac I and Xba I cloning sites of the phagemid as described in Figure 7.

Figure 10 illustrates an ethidium bromide stained agarose gel. After PCR amplification from human cloned DNA of heavy chain alone (HC), light chain alone (LC), and the heavy/light dicistronic DNA molecule (H/L), DNA samples were electrophoresed. The expected sizes of the HC, LC, and H/L products visualized on the gel were approximately 730, 690, and 1,390 base pairs, respectively.

Figure 11 illustrates an autoradiogram showing signals obtained from human phage clones. Approximately 100 lambda phage were spotted onto E. coli lawns, creating plaques that were overlaid with nitrocellulose filters previously soaked in 10 mM isopropylbeta-D-thiogalactopyranoside (IPTG) to induce Fab expression. Following overnight incubation, the filters were reacted with ^{125}I-tetanus toxoid probe. After washing, the filters were exposed to X-ray film. The column on the right represents the parental clones that were selected from a combinatorial library. Mullinax et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990). The column on the left represents clones that were generated by amplifying, the combinatorial lambda clone DNA with the V_H and C_L' outside primers, C_H1' and V_L inside primers, followed by recloning in

the modified ImmunoZAP H vector. Clone 7G1 is a negative control which expresses an Fab that does not react with tetanus toxoid. Clones 10C1 and 6C1 both produce Fabs that react with tetanus toxoid. IZ H is the modified heavy chain ImmunoZAP H vector without an insert.

Figure 12 illustrates the major features of the bacterial expression vector lambda ZAP H/L (ImmunoZAP H/L) (combined V_H- and V_L-expression vector). The ImmunoZAP H/L vector is created from the heavy and light chain libraries by fusing the vectors at the Eco R1 site. DNA is purified from the light chain library and restriction digested with Mlu 1 and Eco R1. This cleaves the DNA from the left arm of the vector into several pieces while leaving the right arm with the light chain inserts intact. DNA is purified from the heavy chain libraries and restriction digested with Hind III and Eco R1. This cleaves the DNA from the right arm of the vector into several pieces while leaving the left arm with the heavy chain inserts intact. The intact left arm of the heavy chain vector containing the heavy chain inserts and right arm of the light chain vector containing the light chain inserts are then mixed and ligated at the common Eco R1 restriction site.

Detailed Description of the Invention

A. Definitions

Nucleotide: A monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3'

or 5' position of the pentose it is referred to as a nucleotide.

Base Pair (bp): A partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

Nucleic Acid: A polymer of nucleotides, either single or double stranded.

Gene: A nucleic acid whose nucleotide sequence codes for an RNA or polypeptide. A gene can be either RNA or DNA.

Complementary Bases: Nucleotides that normally pair up when DNA or RNA adopts a double stranded configuration.

Complementary Nucleotide Sequence: A sequence of nucleotides in a single-stranded molecule of DNA or RNA that is sufficiently complementary to that on another single strand to specifically hybridize to it with consequent hydrogen bonding.

Conserved: A nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-randomly hybridizes to an exact complement of the preselected sequence.

Hybridization: The pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establishment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotides that can be competitively inhibited.

Nucleotide Analog: A purine or pyrimidine nucleotide that differs structurally from A, T, G, C, or U, but is sufficiently similar to substitute for the normal nucleotide in a nucleic acid molecule.

DNA Homolog: Is a nucleic acid having a preselected conserved nucleotide sequence and a sequence coding for a receptor capable of binding a preselected ligand.

Receptor: A receptor is a molecule, such as a protein, glycoprotein and the like, that can specifically (non-randomly) bind to another molecule.

Antibody: The term antibody in its various grammatical forms is used herein to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v).

Antibody Combining Site: An antibody combining site is that structural portion of an antibody molecule comprised of a heavy and light chain variable and hypervariable regions that specifically binds (immunoreacts with) an antigen. The term immunoreact in its various forms means specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or a portion thereof.

Monoclonal Antibody: The phrase monoclonal antibody in its various grammatical forms refers to a population of antibody molecules that contains only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore

contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen, e.g., a bispecific monoclonal antibody.

Upstream: In the direction opposite to the direction of DNA transcription, and therefore going from 5' to 3' on the non-coding strand, or 3' to 5' on the mRNA.

Downstream: Further along a DNA sequence in the direction of sequence transcription or read out, that is traveling in a 3'- to 5'-direction along the non-coding strand of the DNA or 5'- to 3'-direction along the RNA transcript.

Cistron: Sequence of nucleotides in a DNA molecule coding for an amino acid residue sequence.

Stop Codon: Any of three codons that do not code for an amino acid, but instead cause termination of protein synthesis. They are UAG, UAA and UGA. Also referred to as a nonsense or termination codon.

Leader Polypeptide: A short length of amino acid sequence at the amino end of a protein, which carries or directs the protein through the inner membrane and so ensures its eventual secretion into the periplasmic space and perhaps beyond. The leader sequence peptide is commonly removed before the protein becomes active.

Reading Frame: Particular sequence of contiguous nucleotide triplets (codons) employed in translation. The reading frame depends on the location of the translation initiation codon.

Inside Primer: An inside primer is a polynucleotide that has a priming region located at the 3' terminus of the primer which typically consists

of 15 to 30 nucleotide bases. The 3' terminal-priming portion is capable of acting as a primer to catalyze nucleic acid synthesis. The 5'-terminal priming portion comprises a non-priming portion.

Outside Primer: An outside primer comprises a 3'-terminal priming portion and a portion that may define an endonuclease restriction site which is typically located in a 5'-terminal non-priming portion of the outside primer.

B. Methods

The present invention contemplates a method of isolating from a repertoire of conserved genes a pair of genes coding for a dimeric receptor having a preselected activity. Preferably, the receptor will be a heterodimeric polypeptide capable of binding a ligand, such as an antibody molecule or immunologically active portion thereof, a cellular receptor, or a cellular adhesion protein coded for by one of the members of a family of conserved genes, i.e., genes containing a conserved nucleotide sequence of at least about 10 nucleotides in length.

Exemplary conserved gene families encoding different polypeptide claims of a dimeric receptor are those coding for immunoglobulins, major histocompatibility complex antigens of class I or II, lymphocyte receptors, integrins and the like.

A gene can be identified as belonging to a repertoire of conserved genes using several methods. For example, an isolated gene may be used as a hybridization probe under low stringency conditions to detect other members of the repertoire of conserved genes present in genomic DNA using the methods described by Southern, J. Mol. Biol., 98:503 (1975). If the gene used as a hybridization probe hybridizes

to multiple restriction endonuclease fragments of the genome, that gene is a member of a repertoire of conserved genes.

Immunoglobulins

The immunoglobulins, or antibody molecules, are a large family of molecules that include several types of molecules, such as IgD, IgG, IgA, IgM and IgE. The antibody molecule is typically comprised of two heavy (H) and light (L) chains with both a variable (V) and constant (C) region present on each chain as shown in Figure 1. Schematic diagrams of human IgG heavy chain and human kappa light chain are shown in Figures 2A and 2B, respectively. Several different regions of an immunoglobulin contain conserved sequences useful for isolating an immunoglobulin repertoire. Extensive amino acid and nucleic acid sequence data displaying exemplary conserved sequences is compiled for immunoglobulin molecules by Kabat et al., in Sequences of Proteins of Immunological Interest, National Institutes of Health, Bethesda, MD, 1987.

The C region of the H chain defines the particular immunoglobulin type. Therefore the selection of conserved sequences as defined herein from the C region of the H chain results in the preparation of a repertoire of immunoglobulin genes having members of the immunoglobulin type of the selected C region.

The V region of the H or L chain typically comprises four framework (FR) regions each containing relatively lower degrees of variability that includes lengths of conserved sequences. The use of conserved sequences from the FR1 and FR4 (J region) framework regions of the V_H chain is a preferred exemplary

embodiment and is described herein in the Examples. Framework regions are typically conserved across several or all immunoglobulin types and thus conserved sequences contained therein are particularly suited for preparing repertoires having several immunoglobulin types.

Major Histocompatibility Complex

The major histocompatibility complex (MHC) is a large genetic locus that encodes an extensive family of proteins that include several classes of molecules referred to as class I, class II or class III MHC molecules. Paul et al., in Fundamental Immunology, Raven Press, NY, pp. 303-378 (1984).

Class I MHC molecules are a polymorphic group of transplantation antigens representing a conserved family in which the antigen is comprised of a heavy chain and a non-MHC encoded light chain. The heavy chain includes several regions, termed the N, C1, C2, membrane and cytoplasmic regions. Conserved sequences useful in the present invention are found primarily in the N, C1 and C2 regions and are identified as continuous sequences of "invariant residues" in Kabat et al., supra.

Class II MHC molecules comprise a conserved family of polymorphic antigens that participate in immune responsiveness and are comprised of an alpha and a beta chain. The genes coding for the alpha and beta chain each include several regions that contain conserved sequences suitable for producing MHC class II alpha or beta chain repertoires. Exemplary conserved nucleotide sequences include those coding for amino acid residues 26-30 of the A1 region, residues 161-170 of the A2 region and residues 195-206 of the membrane region, all of the alpha chain.

Conserved sequences are also present in the B1, B2 and membrane regions of the beta chain at nucleotide sequences coding for amino acid residues 41-45, 150-162 and 200-209, respectively.

Lymphocyte Receptors and Cell Surface Antigens

Lymphocytes contain several families of proteins on their cell surfaces including the T-cell receptor, Thy-1 antigen and numerous T-cell surface antigens including the antigens defined by the monoclonal antibodies OKT4 (leu3), OKUT5/8 (leu2), OKUT3, OKUT1 (leu1), OKT 11 (leu5) OKT6 and OKT9. Paul, supra at pp. 458-479.

The T-cell receptor is a term used for a family of antigen binding molecules found on the surface of T-cells. The T-cell receptor as a family exhibits polymorphic binding specificity similar to immunoglobulins in its diversity. The mature T-cell receptor is comprised of alpha and beta chains each having a variable (V) and constant (C) region. The similarities that the T-cell receptor has to immunoglobulins in genetic organization and function shows that T-cell receptor contains regions of conserved sequence. Lai et al., Nature, 331:543-546 (1988).

Exemplary conserved sequences include those coding for amino acid residues 84-90 of alpha chain, amino acid residues 107-115 of beta chain, and amino acid residues 91-95 and 111-116 of the gamma chain. Kabat et al., supra, p. 279.

Integrins And Adhesions

Adhesive proteins involved in cell attachment are members of a large family of related proteins termed integrins. Integrins are heterodimers

comprised of a beta and an alpha subunit. Members of the integrin family include the cell surface glycoproteins platelet receptor GpIIb-IIIa, vitronectin, receptor (VnR) fibronectin receptor (FnR) and the leukocyte adhesion receptors LFA-1, Mac-1, Mo-1 and 60.3. Rouslahti et al., Science, 238:491-497 (1987). Nucleic acid and protein sequence data demonstrates regions of conserved sequences exist in the members of these families, particularly between the beta chain of GpIIb-IIIa VnR and FnR, and between the alpha subunit of VnR, Mac-1, LFA-1, FnR and GpIIb-IIIa. Suzuki et al., Proc. Natl. Acad. Sci. USA, 83:8614-8618, 1986; Ginsberg et al., J. Biol. Chem., 262:5437-5440, 1987.

Fusion PCR

In the present invention, fusion PCR is used to generate two PCR-amplified DNA fragments, each of which have one of their ends modified by directed mispriming so that those ends share regions of complementarity, i.e., cohesive termini. When the two fragments are mixed, denatured and reannealed in a PCR cycle, the cohesive termini on two strands hybridize to form an "overlapping" DNA duplex that is internally primed. The subsequent PCR cycle primer-extends the non-overlapping regions to form a hybrid DNA molecule that is dicistronic. See Figure 4.

PCR amplification methods are described in detail in U.S. Patent Nos. 4,683,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including "PCR Technology: Principles and Applications for DNA Amplification", H. Erlich, ed., Stockton Press, New York (1989); and "PCR Protocols: A Guide to Methods and Applications", Innis et al., eds., Academic Press, San Diego, California (1990).

Cloning From Gene Repertoires

The following discussion illustrates the method of the present invention applied to isolating a pair of V_H and V_L genes from the immunoglobulin gene repertoire. This discussion is not to be taken as limiting, but rather as illustrating application of principles that can be used to operatively link and isolate a functionally similar pair of genes. The illustrated method can be used with any family of conserved genes coding for functionally related dimeric receptors, whether obtained directly from a natural source, such naive or in vivo immunized cells, or from cells or one or more genes that have been treated or mutagenized in vitro. Generally, the method, combines the following elements:

1. Producing V_H and V_L gene repertoires.

2. Preparing sets of outside and inside polynucleotide primers for cloning polynucleotide segments containing immunoglobulin V_H and V_L region genes.

3. Preparing a library containing a plurality of different dicistronic DNA molecules, each containing a V_H and a V_L gene from the respective repertoires.

4. Expressing the dicistronic DNA molecules in suitable host cells.

5. Screening the receptors formed by the polypeptides expressed by the dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecule identified by the screening process.

In one method of producing a library of dicistronic DNA molecules containing upstream and downstream cistrons, first and second PCR

amplification products are produced using respective first and second PCR primer pairs. The first PCR primer pair comprises a first polypeptide outside primer and a first polypeptide inside primer. Similarly, the second PCR primer pair comprises a second polypeptide outside primer and a second polypeptide inside primer. The first and second polypeptide inside primers contain complementary 5'-terminal sequences that allow their DNA complements to hybridize and form an internally-primed duplex having 3'-overhanging termini. The internally-primed duplex is then subjected to primer extension reaction conditions to produce a double stranded, dicistronic DNA having substantially blunt or blunt ends. The dicistronic DNA is then PCR amplified using the outside primers as a PCR primer pair.

A dicistronic DNA molecule of this invention contains two amino acid residue-coding sequences on the same strand separated by at least one stop codon and at least one signal sequence necessary for translation of the downstream cistron, such as a translation initiation codon, ribosome binding site, and the like. Thus, the upstream and downstream cistrons of the dicistronic DNA molecule are operatively linked by a cistronic bridge. The cistronic bridge contains the genetic elements necessary to terminate translation of the upstream cistron and initiate translation of the downstream cistron. For instance, the coding strand of the bridge codes for one or more stop codons, preferably two, in the same translational reading frame as the upstream cistron. The cistronic bridge coding strand preferably also encodes a ribosome binding site for the downstream cistron located downstream from the upstream cistron's stop codon(s). Typically, the

coding strand of the cistronic bridge will also encode a leader polypeptide segment in the same translational reading frame as the downstream cistron. When present, the nucleotide base sequence encoding the leader usually begins with an initiation codon located within an operative distance, i.e., is operatively linked, to the ribosome binding site.

A receptor produced by the present invention assumes a conformation having a binding site specific for, as evidenced by its ability to be competitively inhibited, a preselected or predetermined ligand such as an antigen, enzymatic substrate and the like. In one embodiment, a receptor of this invention is a ligand binding heterodimeric polypeptide that forms an antigen binding site which specifically binds to a preselected antigen to form a complex having a sufficiently strong binding between the antigen and the binding site for the complex to be isolated. When the receptor is an antigen binding polypeptide its affinity or avidity is generally greater than 10^5 M^{-1} more usually greater than 10^6 M^{-1} and preferably greater than 10^8 M^{-1}.

In another embodiment, a receptor of the subject invention binds a substrate and catalyzes the formation of a product from the substrate. While the topology of the ligand binding site of a catalytic receptor is probably more important for its preselected activity than its affinity (association constant or pKa) for the substrate, the subject catalytic receptors have an association constant for the preselected substrate generally greater than 10^3 M^{-1}, more usually greater than 10^5 M^{-1} or 10^6 M^{-1} and preferably greater than 10^7 M^{-1}.

Preferably the receptor produced by the subject invention is heterodimeric and is therefore

normally comprised of two different polypeptide chains, which together assume a conformation having a binding affinity, or association constant for the preselected ligand that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. One or both of the different polypeptide chains is derived from the variable region of the light and heavy chains of an immunoglobulin. Typically, polypeptides comprising the light (V_L) and heavy (V_H) variable regions are employed together for binding the preselected ligand.

A receptor produced by the subject invention can be comprised of active monomers V_H and V_L ligand binding polypeptides produced by the present invention can be advantageously combined in the heterodimer to modulate the activity of either or to produce an activity unique to the heterodimer.

The individual ligand polypeptides will be referred to as V_H and V_L and the heterodimer will be referred to as a F_V. However, it should be understood that a V_H may contain in addition to the V_H, substantially all or a portion of the heavy chain constant region. Similarly, a V_L may contain, in addition to the V_L, substantially all or a portion of the light chain constant region. A heterodimer comprised of a V_H containing a portion of the heavy chain constant region and a V_L containing substantially all of the light chain constant region is termed a Fab fragment. The production of Fab can be advantageous in some situations because the additional constant region sequences contained in a Fab as compared to a F_V can stabilize the V_H and V_L interaction. Such stabilization can cause the Fab to have higher affinity for antigen. In addition the Fab

is more commonly used in the art and thus there are more commercial antibodies available to specifically recognize a Fab in screening procedures.

The individual V_H and V_L polypeptides can be produced in lengths equal to or substantially equal to their naturally occurring lengths. See Figure 2. However, in preferred embodiments, the V_H and V_L polypeptides will generally have fewer than 125 amino acid residues, more usually fewer than about 120 amino acid residues, while normally having greater than 60 amino acid residues, usually greater than about 95 amino acid residues, more usually greater than about 100 amino acid residues. Preferably, the V_H will be from about 110 to about 125 amino acid residues in length while V_L will be from about 95 to about 115 amino acid residues in length.

The amino acid residue sequences will vary widely, depending upon the particular idiotype involved. Usually, there will be at least two cysteines separated by from about 60 to 75 amino acid residues and joined by a disulfide bond. The polypeptides produced by the subject invention will normally be substantial copies of idiotypes of the variable regions of the heavy and/or light chains of immunoglobulins, but in some situations a polypeptide may contain random mutations in amino acid residue sequences in order to advantageously improve the desired activity.

In some situations, it is desirable to provide for covalent cross linking of the V_H and V_L polypeptides, which can be accomplished by providing cysteine resides at the carboxyl termini. The polypeptide will normally be prepared free of the immunoglobulin constant regions, however a small portion of the J region may be included as a result of

the advantageous selection of DNA synthesis primers. The D region will normally be included in the transcript of the V_H.

Typically the C terminus region of the V_H and V_L polypeptides will have a greater variety of sequences than the N terminus and, based on the present strategy, can be further modified to permit a variation of the normally occurring V_H and V_L chains. A synthetic polynucleotide can be employed to vary one or more amino acid in a hypervariable region.

1. Producing A Gene Repertoire

A gene repertoire useful in practicing the present invention contains at least 10^3, preferably at least 10^4, more preferably at least 10^5, and most preferably at least 10^7 different conserved genes. Methods for evaluating the diversity of a repertoire of conserved genes is well known to one skilled in the art.

Various well known methods can be employed to produce a useful gene repertoire. For instance, V_H and V_L gene repertoires can be produced by isolating V_H- and V_L-coding mRNA from a heterogeneous population of antibody producing cells, i.e., B lymphocytes (B cells), preferably rearranged B cells such as those found in the circulation or spleen of a vertebrate. Rearranged B cells are those in which immunoglobulin gene translocation, i.e., rearrangement, has occurred as evidenced by the presence in the cell of mRNA with the immunoglobulin gene V, D and J region transcripts adjacently located thereon. Typically, the B cells are collected in a 1-100 ml sample of blood which usually contains 10^6 B cells/ml.

In some cases, it is desirable to bias a repertoire for a preselected activity, such as by

using as a source of nucleic acid cells (source cells) from vertebrates in any one of various stages of age, health and immune response. For example, repeated immunization of a healthy animal prior to collecting rearranged B cells results in obtaining a repertoire enriched for genetic material producing a receptor of high affinity. Mullinax et al., Proc. Natl. Acad. Sci. USA, 87:8095-8099 (1990). Conversely, collecting rearranged B cells from a healthy animal whose immune system has not been recently challenged results in producing a repertoire that is not biased towards the production of high affinity V_H and/or V_L polypeptides.

It should be noted the greater the genetic heterogeneity of the population of cells for which the nucleic acids are obtained, the greater the diversity of the immunological repertoire (comprising V_H- and V_L-coding genes) that will be made available for screening according to the method of the present invention. Thus, cells from different individuals, particularly those having an immunologically significant age difference, and cells from individuals of different strains, races or species can be advantageously combined to increase the heterogeneity (diversity) of a repertoire.

Thus, in one preferred embodiment, the source cells are obtained from a vertebrate, preferably a mammal, which has been immunized or partially immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally. Antibody titer in the animal can be monitored to determine the stage of immunization desired, which stage corresponds to the amount of enrichment or biasing of the repertoire desired. Partially immunized animals typically receive only one

immunization and cells are collected from those animals shortly after a response is detected. Fully immunized animals display a peak titer, which is achieved with one or more repeated injections of the antigen into the host mammal, normally at 2 to 3 week intervals. Usually three to five days after the last challenge, the spleen is removed and the genetic repertoire of the spleenocytes, about 90% of which are rearranged B cells, is isolated using standard procedures. See, Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, NY. Nucleic acids coding for V_H and V_L polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells.

Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned as a diverse population are well known in the art. See for example Herrmann et al., Methods In Enzymol., 152:180-183, (1987); Frischauf, Methods In Enzymol., 152:183-190 (1987); Frischauf, Methods In Enzymol., 152:190-199 (1987); and DiLella et al., Methods In Enzymol., 152:199-212 (1987). (The teachings of the references cited herein are hereby incorporated by reference.)

The desired gene repertoire can be isolated from either genomic material containing the gene expressing the variable region or the messenger RNA (mRNA) which represents a transcript of the variable region. The difficulty in using the genomic DNA from other than non-rearranged B lymphocytes is in juxtaposing the sequences coding for the variable region, where the sequences are separated by introns. The DNA fragment(s) containing the proper exons must be isolated, the introns excised, and the exons then

spliced in the proper order and in the proper orientation. For the most part, this will be difficult, so that the alternative technique employing rearranged B cells will be the method of choice because the V, D and J immunoglobulin gene regions have translocated to become adjacent, so that the sequence is continuous (free of introns) for the entire variable regions.

Where mRNA is utilized the cells will be lysed under RNase inhibiting conditions. In one embodiment, the first step is to isolate the total cellular mRNA. Poly A+ mRNA can then be selected by hybridization to an oligo-dT cellulose column. The presence of mRNAs coding for the heavy and/or light chain polypeptides can then be assayed by hybridization with DNA single strands of the appropriate genes. Conveniently, the sequences coding for the constant portion of the V_H and V_L can be used as polynucleotide probes, which sequences can be obtained from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender, eds., Vol. 3, Plenum Publishing Corporation, NY, (1981), pages 157-188; and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, MD, (1987). In preferred embodiments, the preparation containing the total cellular mRNA is first enriched for the presence of V_H and/or V_L coding mRNA. Enrichment is typically accomplished by subjecting the total mRNA preparation or partially purified mRNA product thereof to a primer extension reaction employing a polynucleotide synthesis primer of the present invention. Exemplary methods for producing V_H and V_L gene repertoires are described in PCT Application No. PCT/US 90/02836 (International Publication No. WO 90/14430).

In preferred embodiments, isolated B cells are immunized in vitro against a preselected antigen. In vitro immunization is defined as the clonal expansion of epitope-specific B cells in culture, in response to antigen stimulation. The end result is to increase the frequency of antigen-specific B cells in the immunoglobulin repertoire, and thereby decrease the number of clones in an expression library that must be screened to identify a clone expressing an antibody of the desired specificity. The advantage of in vitro immunization is that human monoclonal antibodies can be generated against a limitless number of therapeutically valuable antigens, including toxic or weak immunogens. For example, antibodies specific for the polymorphic determinants of tumor-associated antigens, rheumatoid factors, and histocompatibility antigens can be produced, which can not be elicited in immunized animals. In addition, it may be possible to generate immune responses which are normally suppressed in vivo.

In vitro immunization can be used to give rise to either a primary or secondary immune response. A primary immune response, resulting from first time exposure of a B cell to an antigen, results in clonal expansion of epitope-specific cells and the secretion of IgM antibodies with low to moderate apparent affinity constants (10^6-$10^8 M^{-1}$). Primary immunization of human splenic and tonsillar lymphocytes in culture can be used to produce monoclonal antibodies against a variety of antigens, including cells, peptides, macromolecules, haptens, and tumor-associated antigens. Memory B cells from immunized donors can also be stimulated in culture to give rise to a secondary immune response characterized by clonal expansion and the production of high affinity

antibodies (>10^9 M^{-1}) of the IgG isotype, particularly against viral antigens by clonally expanding sensitized lymphocytes derived from seropositive individuals.

In one embodiment, peripheral blood lymphocytes are depleted of various cytolytic cells that appear to down-modulate antigen-specific B cell activation. When lysosome-rich subpopulations (natural killer cells, cytotoxic and suppressor T cells, monocytes) are first removed by treatment with the lysosmotropic methyl ester of leucine, the remaining cells (including B cells, T helper cells, accessory cells) respond antigen-specifically during in vitro immunization. The lymphokine requirements for inducing antibody production in culture are satisfied by a culture supernatant from activated, irradiated T cells.

In addition to in vitro immunization, cell panning (immunoaffinity absorption) can be used to further increase the frequency of antigen-specific B cells. Techniques for selecting B cell subpopulations via solid-phase antigen binding are well established. Panning conditions can be optimized to selectively enrich for B cells which bind with high affinity to a variety of antigens, including cell surface proteins. Panning can be used alone, or in combination with in vitro immunization to increase the frequency of antigen-specific cells above the levels which can be obtained with either technique alone. Immunoglobulin expression libraries constructed from enriched populations of B cells are biased in favor of antigen-specific antibody clones, and thus, enabling identification of clones with the desired specificities from smaller, less complex libraries.

2. Preparation Of Polynucleotide Primers

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than 3. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be in double stranded form. If double stranded, the primer is first treated to separate it from its complementary strand before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on may factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally require

cooler temperatures to form sufficiently stable hybrid complexes with template.

The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently complementary to non-randomly hybridize with its respective template strand. Therefore, the primer sequence may or may not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such non-complementary fragments typically code for an endonuclease restriction site. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarily with the sequence of the strand to be synthesized or amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucleic Acids Research, 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

When a primer containing a DNA-dependent RNA polymerase promoter is used the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing

agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA polynucleotide.

Primers may also contain a template sequence or replication initiation site for a RNA-directed RNA polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al., Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numbers of RNA strands from a small number of template RNA strands that contain a template sequence or replication initiation site. These polymerases typically give a one million-fold amplification of the template strand as has been described by Kramer et al., J. Mol. Biol., 89:719-736 (1974).

The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester or phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patent No. 4,356,270; and Brown et al., Meth. Enzymol., 68:109, (1979).

The choice of a primer's nucleotide sequence depends on factors such as the distance on the nucleic acid from the region coding for the desired receptor, its hybridization site on the nucleic acid relative to any second primer to be used, the number of genes in the repertoire it is to hybridize to, and the like.

(a) Primers for Producing Gene Repertoires

V_H and V_L gene repertoires can be separately prepared prior to their utilization in the present invention. Repertoire preparation is typically accomplished by primer extension, preferably

by primer extension in a PCR format.

To produce a repertoire of V_H-coding DNA homologs by primer extension, the nucleotide sequence of a primer is selected to hybridize with a plurality of immunoglobulin heavy chain genes at a site substantially adjacent to the V_H-coding region so that a nucleotide sequence coding for a functional (capable of binding) polypeptide is obtained. To hybridize to a plurality of different V_H-coding nucleic acid strands, the primer must be a substantial complement of a nucleotide sequence conserved among the different strands. Such sites include nucleotide sequences in the constant region, any of the variable region framework regions, preferably the third framework region, leader region, promoter region, J region and the like.

If the repertoires V_H-coding and V_L-coding DNA homologs are to be produced by polymerase chain reaction (PCR) amplification, two primers, i.e., a PCR primer pair, must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the nonsense (minus or complementary) strand and hybridizes to a nucleotide sequence conserved among V_H (plus or coding) strands within the repertoire. To produce V_H coding DNA homologs, first primers are therefore chosen to hybridize to (i.e. be complementary to) conserved regions within the J region, CH1 region, hinge region, CH2 region, or CH3 region of immunoglobulin genes and the like. To produce a V_L coding DNA homolog, first primers are chosen to hybridize with (i.e. be complementary to) a conserved region within the J region or constant region of immunoglobulin light chain genes and the like. Second primers become part of the coding (plus) strand and hybridize to a nucleotide sequence

conserved among minus strands. To produce the V_H-coding DNA homologs, second primers are therefore chosen to hybridize with a conserved nucleotide sequence at the 5' end of the V_H-coding immunoglobulin gene such as in that area coding for the leader or first framework region. It should be noted that in the amplification of both V_H- and V_L-coding DNA homologs the conserved 5' nucleotide sequence of the second primer can be complementary to a sequence exogenously added using terminal deoxynucleotidyl transferase as described by Loh et al., Sci. Vol 243:217-220 (1989). One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site. The site can be heterologous to the immunoglobulin gene being amplified and typically appears at or near the 5' end of the primer.

(b) Inside and Outside Primers

In one embodiment, the present invention utilizes a set of polynucleotides that form inside primers comprised of an upstream inside primer and a downstream inside primer. Each of the inside primers has a priming region located at the 3'-terminus of the primer. The priming region is typically the 3'-most (3'-terminal) 15 to 30 nucleotide bases. The 3'-terminal priming portion of each inside primer is capable of acting as a primer to catalyze nucleic acid synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the inside primers is further characterized by the presence of a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to repertoire template.

In fusion PCR, each inside primer works in combination with an outside primer to amplify a target nucleic acid sequence. The choice of PCR primer pairs for use in fusion PCR as described herein is governed by the same considerations as previously discussed for choosing PCR primer pairs useful in producing gene repertoires. That is, the primers have a nucleotide sequence that is complementary to a sequence conserved in the repertoire. Useful V_L and V_H inside priming sequences are shown in Tables 1 and 2, respectively, below.

Table 1

3' Priming Portions of Various Inside V_L Primers

Seq. Id. No.	
(1)[1]	5' GTGATGACCCACTCTCC 3'
(2)	5' GTGATGACCCAGTCTCCA 3'
(3)	5' GTTGTGACTCAGGAATCT 3'
(4)	5' GTGTTGACGCAGCCGCCC 3'
(5)	5' GTGCTCACCCAGTCTCCA 3'
(6)	5' CAGATGACCCAGTCTCCA 3'
(7)	5' GTGATGACCCAGACTCCA 3'
(8)	5' GTCATGACCCAGTCTCCA 3'
(9)	5' TTGATGACCCAAACTCAA 3'
(10)	5' GTGATAACCCAGGATGAA 3'

[1] Nucleotide sequences 1-10 are unique 5' primers for the amplification of kappa light chain variable regions.

Table 2

3' Priming Portions of Various Inside V_H Primers

	Seq. Id. No.	
	(11)[1]	5' ACAAGATTTGGGCTC 3'
	(12)[2]	5' TGGGGTTTTGAGCTC 3'
	(13)[3]	5' GAGACAGTGACCGGGTTCCTTGGCCCCA 3
	(14)[4]	5' TGGAATGGGCACATGCAG 3'
	(15)[5]	5' TTATCATTTACCCGGAGA 3'
	(16)[6]	5' AACGGTAACAGTGGTGCCTTGGCCCCA 3'
	(17)[7]	5' ACAATCCCTGGGCACAAT 3'
	(18)[8]	5' CACCTTGGTGCTGCTGGC 3'
	(19)[9]	5' ACAACCACAATCCCTGGGCACAATTTT 3'
	(20)[10]	5' ACAATCCCTGGGCACAAT 3'
	(21)[11]	5' GAGTTCACTAGTTGGGCACGGTGGGCA 3'

[1] Unique 3' primer for human IgG1, 2, 3 and 4 Fd.

[2] Unique 3' primer for human V_H amplification.

[3] 3' primer for amplifying human heavy chain variable regions.

[4] 3' primer for amplifying the Fd region of mouse IgM.

[5] 3' primer located in the CH3 region of human IgG1 to amplify the entire heavy chain.

[6] Unique 3' primer for amplification of mouse F_V.

[7] Unique 3' primer for amplification of mouse IgG1 Fd.

[8] Unique 3' primer for amplification of V_H including part of the mouse gamma 1 first constant region.

[9] Unique 3' primer for amplification of V_H including part of mouse gamma 1 first constant region and hinge region.

[10] 3' primer for amplifying mouse Fd including part of the mouse IgG first constant region and part of the hinge region.

[11] 3' primer for amplifying human IgG1 Fd including part of the human IgG first constant region and part of the hinge region including the two cysteines which create the disulfide bridge for producing Fab'2 (the primer corresponds to Kabat numbers 241QQ to 247).

A preferred set of inside primers used herein has primers with complementary 5'-terminal non-priming regions, the complementary strands of which are capable of hybridizing to each other to form a duplex with 3' overhangs. The duplex encodes all or part of a double stranded cistronic bridge. That is, if the 3' overhangs of the duplex are filled in with complementary bases so as to define a double stranded DNA extending from the 3'-terminus of one of the inside primers to the 3'-terminus of the other of the inside primers, that double stranded DNA segment forms a sequence of nucleotides that operatively links the upstream and downstream cistrons for polycistronic expression. Thus, while each of the inside primers in a set contains only a portion of the sequence information necessary to form the double stranded cistronic bridge, the two inside primers in combination encode both the plus and minus strands of all or part of the bridge.

For example, one inside upstream primer can have a sequence that forms a portion of the plus strand of the bridge, and the other inside primer encodes the sequence, through complementarity, of the downstream portion of the plus strand.

In a preferred embodiment, the plus strand of the cistronic bridge contains, in the translational reading frame and from an upstream position to a downstream position, sequences coding for (i) at least one stop codon, preferably two, in the same reading frame as the upstream cistron, (ii) a ribosome binding

site, and (iii) a polypeptide leader, the translation initiation codon of which is in the same reading frame as the downstream cistron. The stop codon is present to terminate translation of the upstream cistron. The ribosome binding site is present to initiate translation of the downstream cistron from the polycistronic mRNA.

The predicted amino acid residue sequences of two pelB gene product variants from Erwinia Carotova are shown in Table 3. Lei, et al., supra., Amino Acid residue sequences for other leaders from E. coli useful in this invention are also listed in Table 3. Oliver, In Neidhart, F. C. (ed.), Escherichia coli and Salmonella Typhimurium, American Society for Microbiology, Washington, D.C., 1:56-69 (1987). These regions for the heavy chain are contained in the modified ImmunoZAP H expression vector. Mullinax, et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990).

Table 3

Leader Sequences

Seq. Id. No.	Type	Amino Acid Residue Sequence
(22)	pelB[1]	MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeu LeuLeuAlaAlaGlnProAlaGlnProAlaMetAla
(23)	pelB[2]	MetLysSerLeuIleThrProIleAlaAlaGlyLeuLeu LeuAlaPheSerGlnTyrSerLeuAla
(24)	MalE[3]	MetLysIleLysThrGlyAlaArgIleLeuAlaLeuSer AlaLeuThrThrMetMetPheSerAlaSerAlaLeuAla LysIle
(25)	OmpF[3]	MetMetLysArgAsnIleLeuAlaValIleValProAla LeuLeuValAlaGlyThrAlaAsnAlaAlaGlu
(26)	PhoA[3]	MetLysGlnSerThrIleAlaLeuAlaLeuLeuProLeu LeuPheThrProValThrLysAlaArgThr

(27)	Bla^3	MetSerIleGlnHisPheArgValAlaLeuIleProPhe PheAlaAlaPheCysLeuProValPheAlaHisPro
(28)	$LamB^3$	MetMetIleThrLeuArgLysLeuProLeuAlaValAla ValAlaAlaGlyValMetSerAlaGlnAlaMetAlaVal Asp
(29)	Lpp^3	MetLysAlaThrLysLeuValLeuGlyAlaValIleLeu GlySerThrLeuLeuAlaGlyCysSer

[1] pelB from Erwinia carotovora gene

[2] pelB from Erwinia carotovora EC 16 gene

[3] leader sequences from E. coli

To achieve high levels of gene expression in E. coli, it is necessary to use not only strong promoters to generate large quantities of mRNA, but also ribosome binding sites to ensure that the mRNA is efficiently translated. In E. coli, the ribosome binding site includes an initiation codon (AUG) and a sequence 3-9 nucleotides long located 3 11 nucleotides upstream from the initiation codon [Shine et al., Nature, 254:34 (1975). The sequence, AGGAGGU, which is called the Shine-Dalgarno (SD) sequence, is complementary to the 3' end of E. coli 16S mRNA. Binding of the ribosome to mRNA and the sequence at the 3' end of the mRNA can be affected by several factors:

(i) The degree of complementarity between the SD sequence and 3' end of the 16S tRNA.

(ii) The spacing and possibly the DNA sequence lying between the SD sequence and the AUG [Roberts et al., Proc. Natl. Acad. Sci. USA, 76:760 (1979a); Roberts et al., Proc. Natl. Acad. Sci. USA, 76:5596 (1979b); Guarente et al., Science, 209:1428 (1980); and Guarente et al., Cell, 20:543 (1980).] Optimization is achieved by measuring the level of

expression of genes in plasmids in which this spacing is systematically altered. Comparison of different mRNAs shows that there are statistically preferred sequences from positions -20 to +13 (where the A of the AUG is position 0) [Gold et al., Annu. Rev. Microbiol., 35:365 (1981)]. Leader sequences have been shown to influence translation dramatically (Roberts et al., 1979 a, b supra).

(iii) The nucleotide sequence following the AUG, which affects ribosome binding [Taniguchi et al., J. Mol. Biol., 118:533 (1978)].

Useful ribosome binding sites are shown in Table 4 below.

Table 4

Ribosome Binding Sites*

	Seq. Id. No.		
1.	(30)	5'	AAUCUUGGAGGCUUUUUUAUGGUUCGUUCU
2.	(31)	5'	UAACUAAGGAUGAAAUGCAUGUCUAAGACA
3.	(32)	5'	UCCUAGGAGGUUUGACCUAUGCGAGCUUUU
4.	(33)	5'	AUGUACUAAGGAGGUUGUAUGGAACAACGC

* Sequences of initiation regions for protein synthesis in four phage mRNA molecules are underlined.

AUG = initiation codon (double underlined)

1. = Phage ϕX174 gene-A protein
2. = Phage Qβ replicase
3. = Phage R17 gene-A protein
4. = Phage lambda gene-cro protein

It is preferred that the complementary (overlapping) region of the inside primers and the priming portion of the inside primers have about the same denaturation temperature, Td. The Td of a

sequence can be estimated by the following formula: Td = 4(C+G) + 2(A+T), where C, G, A and T represent the respective number of cytosine, guanine, adenine and thymine bases in the sequence. A Td for the above-identified hybridizing region of about 45-55°C, preferably about 50°C, is preferred. Typically, overlapping regions in the range of about 15 to 20 nucleotides works well in conjunction with priming regions in the range of 15-30 nucleotides.

The set of outside primers forms the termini of the dicistronic DNA molecule. The set of outside primers comprises an upstream outside primer and a downstream outside primer. The outside primers each comprise a 3'-terminal priming portion, and preferably a portion that defines an endonuclease restriction site. When present, the restriction site-defining portion is typically located in a 5'-terminal non-priming portion of the outside primer. The restriction site defined by the upstream outside primer is typically chosen to be one recognized by a restriction enzyme that does not recognize the restriction site defined by the downstream outside primer, the objective being to be able to produce a dicistronic DNA having cohesive termini that are non-complementary to each other and thus allow directional insertion into a vector.

Useful outside primer sequences are shown in Tables 5 and 6 below.

Table 5

Outside V_H Primers

Seq. Id. No.	
(34)[1]	5'AGGTCCAGCTGCTCGAGTCTGG3'
(35)	5'AGGTCCAGCTGCTCGAGTCAGG3'
(36)	5'AGGTCCAGCTTCTCGAGTCTGG3'
(37)	5'AGGTCCAGCTTCTCGAGTCAGG3'
(38)	5'AGGTCCAACTGCTCGAGTCTGG3'
(39)	5'AGGTCCAACTGCTCGAGTCAGG3'
(40)	5'AGGTCCAACTTCTCGAGTCTGG3'
(41)	5'AGGTCCAACTTCTCGAGTCAGG3'
(42)[2]	5'AGGTGCAGCTGCTCGAGTCTGG3'
(43)	5'AGGTGCAGCTGCTCGAGTCGGG3'
(44)	5'AGGTGCAACTGCTCGAGTCTGG3'
(45)	5'AGGTGCAACTGCTCGAGTCGGG3'

[1] Nucleotide sequences 21-28 are unique 5' primers for the amplification of mouse V_H genes.

[2] Nucleotide sequences 29-32 are unique 5' primers for amplification of nucleic acids coding for human variable regions.

Table 6

Outside V_L Primers

Seq. Id. No.	
(46)[1]	5' ACGTCTAGATTCCACCTTGGTCCC 3'
(47)[2]	5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'
(48)[3]	5' GCATTCTAGACTATTAACATTCTGTAGGGGC 3'
(49)[4]	5' GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC 3'
(50)[5]	5' CCGCCGTCTAGAACACTCATTCCTGTTGAAGCT 3'
(51)[6]	5' CCGCCGTCTAGAACATTCTGCAGGAGACAGACT 3'

(52)[7] 5' GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA 3'

(53)[8] 5' GCCGCTCTAGAACACTCATTCCTGTTGAA 3'

(54)[9] 5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'

(55)[10] 5' GCATTCTAGACTATTATGAACATTCTGTAGGGGC 3'

[1] 3' primer for amplifying human kappa chain variable regions.

[2] 3' primer in human kappa light chain constant region.

[3] 3' primer in human lambda light chain constant region.

[4] Unique 3' primer for amplification of kappa light chain variable regions.

[5] Unique 3' primer for mouse kappa light chain amplification including the constant region.

[6] Unique 3' primer for mouse lambda light chain amplification including the constant region.

[7] Unique 3' primer for amplification of kappa light chain.

[8] Unique 3' primer for amplification of mouse kappa light chain.

[9] Unique 3' primer for kappa V_L amplification.

[10] Unique 3' primer for human, mouse and rabbit lambda V_L amplification.

3. Preparing a Dicistronic DNA Molecule Library

The strategy used for cloning the V_H and V_L genes contained within a repertoire will depend, as is well known in the art, on the type, complexity, and purity of the nucleic acids making up the repertoire. Other factors include whether or not the genes are contained in one or a plurality of repertoires and whether or not they are to be amplified and/or mutagenized.

In one embodiment, a library of dicistronic DNA molecules containing upstream and downstream cistrons operatively linked by a cistronic bridge can be produced by the following steps:

(a) Subjecting a repertoire of first polypeptide genes (e.g., V_H-coding genes), to PCR amplification using first outside and first inside primers, i.e., a first PCR primer pair, to form a first primary PCR product.

(b) Subjecting a repertoire of second polypeptide genes (e.g., V_L-coding genes) to PCR amplification using second outside and second inside primers, i.e., a second PCR primer pair, to form a second primary PCR product.

(c) Hybridizing the first and second primary PCR products to form internally (self) primed duplexes, i.e., duplexes having 3'-hybridized and 5'-overhanging termini.

(d) Subjecting the internally-primed duplexes to primer extension reaction conditions to form double stranded duplexes having substantially blunt, preferably blunt, termini and a dicistronic strand containing the upstream and downstream cistrons linked by a cistronic bridge encoded by the inside primers. By "substantially blunt" is meant having no more than about one or two overhanging nucleotides. (Substantially blunt double stranded DNA is sometimes produced by primer overextension by Taq polymerase, usually by the addition of one or two terminal adenine residues.)

The V_H- and V_L-coding gene repertoires are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. A repertoire is

subjected to a PCR reaction by treating (contacting) the repertoire with a PCR primer pair, each member of the pair having a preselected nucleotide sequence. The PCR primer pair is capable of initiating primer extension reactions by hybridizing to nucleotide sequences, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved within the repertoire. The first primer of a PCR primer pair is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid. In addition, the second primer of a PCR primer pair is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand of a nucleic acid, i.e., a strand complementary to a coding strand.

The PCR reaction is performed by mixing the PCR primer pair, preferably a predetermined amount thereof, with the nucleic acids of the repertoire, preferably a predetermined amount thereof, in a PCR buffer to form a PCR reaction admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a PCR reaction product, thereby producing a plurality of different V_H-coding and/or V_L-coding DNA homologs.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of different second primers to form several different primer pairs. Alternatively, an individual pair of first and second primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and

second primers can be combined to increase the diversity of the gene library.

In another strategy, the object is to clone the V_H- and/or V_L-coding genes from a repertoire by providing a polynucleotide complement of the repertoire, such as the anti-sense strand of genomic dsDNA or the polynucleotide produced by subjecting mRNA to a reverse transcriptase reaction. Methods for producing such complements are well known in the art.

The PCR reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, i.e., a PCR buffer, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess (for genomic nucleic acid, usually about 10^6:1 primer:template) of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

In preferred embodiments the ratio of gene molecules and their respective primers is as follows: about 1 x 10^3 V_H gene molecules to about 1 x 10^8 outside V_H primer molecules, about 1 x 10^3 V_H gene molecules, to about 1 x 10^7 inside V_H gene primer molecules, about 1 x 10^3 V_L gene molecules to about 1 x 10^8 outside V_L gene primer molecules, about 1 x 10^4 V_L gene molecules to about 1 x 10^7 V_L gene primer molecules. In more preferred embodiments, 10^4 outside V_H gene primer molecules and 10^3 inside V_H gene primer molecules are used for every V_H gene molecule present in the PCR admixture. Similarly, 10^4 outside V_L gene primer molecules and 10^3 V_L inside gene primer molecules are used for every V_L gene molecule present in the PCR admixture. Thus, there is typically a 10 fold molar excess of outside primer to inside primer.

In the fusion PCR reaction, the gene repertoires are admixed with outside and inside primers, the outside primers being present in excess relative to the inside primers. The initial PCR thermocycles produce intermediate products having complementary termini from each of the first and second gene repertoires. That is, the end of one strand from one primary PCR product is capable of hybridizing with the complementary end from the other primary PCR product. The strands having the overlap at their 3' ends can act as primers for one another, i.e., from an internally primed duplex, and be extended by the polymerase to form the full length final product. The final product is then amplified by the set of outside primers, which act as a third PCR pair when the inside primers have been exhausted, to form a secondary PCR product. Typically the molar ratio of outside primers to inside primers is such that the inside primers are effectively exhausted within about 2 to about 12, preferably about 5, 6 or 7 thermocycles.

The PCR buffer also contains the deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferable for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature

is generally no greater than about 40°C. An exemplary PCR buffer comprises the following: 50 mM KCl; 10 mM Tris-HCl; pH 8.3; 1.5 mM $MgCl_2$; 0.001% (wt/vol) gelatin, 200 μM dATP; 200 μM dTTP; 200 μM dCTP; 200 μM dGTP; and 2.5 units Thermus aquaticus DNA polymerase I (U.S. Patent No. 4,889,818) per 100 microliters of buffer.

The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

The inducing agent also may be a compound or system which will function to accomplish the synthesis of RNA primer extension products, including enzymes. In preferred embodiments, the inducing agent may be a DNA-dependent RNA polymerase such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase. These polymerases produce a complementary RNA polynucleotide. The high turn over rate of the RNA polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes,

ed. P. Boyer, PP. 87-108, Academic Press, New York (1982). Another advantage of T7 RNA polymerase is that mutations can be introduced into the polynucleotide synthesis by replacing a portion of cDNA with one or more mutagenic oligodeoxynucleotides (polynucleotides) and transcribing the partially-mismatched template directly as has been previously described by Joyce et al., Nucleic Acid Research, 17:711-722 (1989). Amplification systems based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, pp 245-252, Academic Press, Inc., San Diego, CA (1990).

If the inducing agent is a DNA-dependent RNA polymerase and therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and the resulting solution is treated as described above.

The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

The first and/or second PCR reactions discussed above can advantageously be used to incorporate into the receptor a preselected epitope useful in immunologically detecting and/or isolating a receptor. This is accomplished by utilizing a first and/or second polynucleotide synthesis primer or expression vector to incorporate a predetermined amino acid residue sequence into the amino acid residue sequence of the receptor.

After producing operatively linked V_H- and V_L-coding DNA homologs for a plurality of different V_H- and V_L-coding genes within the repertoires, the

dicistronic DNA molecules are typically further amplified. While the dicistronic DNA molecules can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the molecules by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a vector. In fact, in preferred strategies, the first and second PCR reactions are performed in the same admixture that is subject to a multiplicity of PCR thermocycles where the outside primers are in molar excess. Preferably the number of PCR thermocycles is at least n+5, wherein n is the number of PCR thermocycles necessary to decrease by a factor of 10, and preferably exhaust, the number of inside primers by consumption in the formation of inside primer-primed products.

PCR is typically carried out by thermocycling i.e., repeatedly increasing and decreasing the temperature of a PCR reaction admixture within a temperature range whose lower limit is about 10°C to about 40°C and whose upper limit is about 90°C to about 100°C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favoring polynucleotide synthesis, denaturation and hybridization.

In preferred embodiments only one pair of first and second primers is used per amplification reaction. The amplification reaction products obtained from a plurality of different amplifications, each using a plurality of different primer pairs, are then combined.

However, the present invention also contemplates DNA homolog production via co-amplification (using two pairs of primers), and

multiplex amplification (using up to about 8, 9 or 10 primer pairs).

A diverse library of dicistronic DNA molecules having upstream and downstream cistrons can also be produced by combining, in a PCR buffer, double stranded V_H and V_L repertoires, V_H and V_L outside primers, and an inside primer having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal inside primer-template (primer-coding) portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion has a nucleotide base sequence homologous (identical) to a portion of the primer extension product of the other of the outside primers. That is, the linking primer has terminal sequences homologous to sequences in both repertoires. The cistronic bridge coding portion codes for, either directly or through complementarily, at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site located downstream from the stop codon, and a polypeptide leader having a translation initiation codon in the same reading frame as the downstream cistron, the initiation codon being located downstream from the ribosome binding site.

The dicistronic DNA molecules containing operatively linked V_H- and V_L-coding DNA homologs produced by PCR amplification are typically in double-stranded form and may have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the dicistronic DNA molecules having restriction sites at or near their termini with one or more appropriate endonucleases results in the production of DNA

molecules having cohesive termini of predetermined specificity.

When individual PCR admixtures contain diverse gene repertoires the present invention produces many non-naturally occurring antibodies, i.e., combinations of V_H and V_L in a heterodimer. To take advantage of the mammalian immune system's capacity to select V_H and V_L combinations, the present invention also contemplates using fusion PCR to operatively link, and thereby recover, naturally occurring V_H and V_L combinations.

In preferred embodiments, a fusion PCR method of the present invention is performed on repertoires comprising a plurality of substantially isolated cells containing genes coding for a heterodimeric receptor. For example, a plurality of PCR admixtures is formed, each of which contains (i) a sample of substantially isolated B lymphocytes from a mammal producing antibody molecules against a preselected antigen, (ii) a PCR buffer, and (iv) either the previously described V_H and V_L PCR primer pairs or the set of outside V_H and V_L PCR primers in combination with the linking primer(s), also as previously described. The plurality of PCR admixtures is then subjected to a multiplicity of PCR thermocycles as described herein.

By "substantially isolated" is meant a sample containing less than about 100 target cells, such as B lymphocytes, T cells, and the like. In preferred embodiments, the plurality of PCR admixtures contain only about one cell. The cells are typically obtained from an individual mammal whose serum contains antibody molecules against the preselected antigen. The collected cells are typically seeded, usually at densities in the range of 0.5 to 100 cells per unit volume, into a plurality of individual PCR vessels,

such as microtiter plate wells and the like. Usually, the plurality of PCR admixtures is in the range of 800 to 1200, and preferably is about 1000, separate admixtures.

Typically, fewer cells are needed in each PCR admixture where the cells are obtained from individuals expressing a high serum antibody titer against the preselected antigen. For example, where B lymphocytes are obtained from an individual having a frequency of circulating B cells producing the antibody molecules of preselected specificity of 1/3000, each of about 800 to 1200 individual PCR admixtures need only contain about one B lymphocyte to result in isolation of the desired antibody. Where the circulating B cell frequency is in the range of 1/500,000, a density of about 100 cells per PCR admixture in each of about 800 to 1200 individual PCR admixtures will be needed before the process will result in isolation of the desired antibody.

In preferred embodiments, the PCR process is used not only to produce a library of dicistronic DNA molecules, but also to induce mutations within the library or to create diversity from a single parental clone and thereby provide a library having a greater heterogeneity. First, it should be noted that the PCR process itself is inherently mutagenic due to a variety of factors well known in the art. Second, in addition to the mutation inducing variations described in the above referenced U.S. Patent No. 4,683,195, other mutation inducing PCR variations can be employed. For example, the PCR reaction admixture, can be formed with different amounts of one or more of the nucleotides to be incorporated into the extension product. Under such conditions, the PCR reaction proceeds to produce nucleotide substitutions within

the extension product as a result of the scarcity of a particular base. Similarly, approximately equal molar amounts of the nucleotides can be incorporated into the initial PCR reaction admixture in an amount to efficiently perform X number of cycles, and then cycling the admixture through a number of cycles in excess of X, such as, for instance, 2X.

Alternatively, mutations can be induced during the PCR reaction by incorporating into the reaction admixture nucleotide derivatives such as inosine, not normally found in the nucleic acids of the repertoire being amplified. During subsequent <u>in vivo</u> amplification, the nucleotide derivative will be replaced with a substitute nucleotide thereby inducing a point mutation.

4. <u>Expressing the Dicistronic DNA Molecules</u>

The dicistronic DNA molecules produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid molecule capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors".

The choice of vector to which a V_H- and V_L-coding DNA homolog is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or

protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. In preferred embodiments, the vector utilized includes a prokaryotic replicon i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra chromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a prokaryotic replicon also include a gene whose expression confers a selective advantage, such as drug resistance, to a bacterial host transformed therewith. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a prokaryotic replicon can also include a prokaryotic promoter capable of directing the expression (transcription and translation) of the V_H- and V_L-coding homologs in a bacterial host cell, such as E. coli transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur.

Promoters contain two highly conserved regions, one located about 10 bp (-10 region on Priberow box) and the other about 35 bp (-35 region) upstream from the point at which transcription starts. These two regions typically determine promoter strength. In addition, the number of nucleotides that separate the conserved sequences is important for efficient promoter function. For example, 16 to 19 nucleotides typically separate the -10 and -35 regions, and changes in that spacing can change the efficiency of a promoter. Promoter sequences

compatible with bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL and pKK223 available from Pharmacia, (Piscataway, NJ).

Promoters useful in this invention include Ptac ϕ 1.1A, ϕ 1.1B and ϕ10, which are recognized by T7 polymerase. See U.S. Patent No. 4,946,786. Useful regulatable promoters include the E. coli lac promoter described in U.S. Patent No. 4,946,786 and the promoters for the temperature sensitive genes in U.S. Patent No. 4,806,471. See also U.S. Patent No. 4,711,845.

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary cohesive termini can be engineered into the dicistronic DNA molecules during the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The dicistronic DNA molecule, and vector if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the vector. The complementary cohesive termini of the vector and the dicistronic DNA molecule are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

The present method produces a diverse population of double stranded DNA expression vectors wherein each vector expresses, under the control of a single promoter, one V_H-coding DNA homolog and one V_L-coding DNA homolog, the diversity of the population being the result of different V_H- and V_L-coding DNA

homolog combinations that occurs during the PCR reaction where both outside and both inside primers are present in effective amounts. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein.

In preferred embodiments, the vector defines a nucleotide sequence coding for a ribosome binding site and a leader, the sequence being located downstream from a promoter and upstream from a sequence coding for a polypeptide leader. In preferred embodiments, the vector contains a selectable marker such that the presence of a dicistronic DNA molecule of this invention inserted into the vector, can be selected. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the V_H- and V_L-coding DNA homologs. When coexpressed within the same organism, a functionally active heterodimeric receptor, such as an F_v, is produced. Cellular hosts into which a V_H- and V_L-coding DNA homolog-containing construct has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferred prokaryotic host cells for library screening, and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eukaryotic host cells include yeast and mammalian

cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al., Proc. Natl. Acad. Sci., USA, 69:2110 (1972); and Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proc. Natl. Acad. Sci., USA, 76:1373-1376 (1979).

5. Screening For Expression of V_H and/or V_L Polypeptides

Successfully transformed cells, i.e., cells containing a dicistronic DNA molecule operatively linked to a vector, can be identified by any suitable well known technique for detecting the binding of a receptor to a ligand or the presence of a polynucleotide coding for the receptor, preferably its active site. Preferred screening assays are those where the binding of ligand by the receptor produces a detectable signal, either directly or indirectly. Such signals include, for example, the production of a complex, formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, cells from a population subjected to transformation with a subject recombinant DNA (rDNA) can be cloned to produce monoclonal colonies. Cells form those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA

using a method such as that described by Southern, J. Mol. Biol., 98:503 (1975) or Berent et al., Biotech., 3:208 (1985).

In addition to directly assaying for the presence of a dicistronic DNA molecule, successful transformation can be confirmed by well known immunological methods, especially when the V_H and/or V_L polypeptides produced contain a preselected epitope. For example, samples of cells suspected of being transformed are assayed for the presence of the preselected epitope using an antibody against the epitope.

Surface Expression

The present invention includes a method for expressing a polypeptide on the outer surface of E. coli. The surface expression of a polypeptide provides a particularly advantageous technique for screening diverse libraries for a polypeptide, such as a receptor, having a pre-selected activity. For example, E. coli expressing a diverse library of Fab fragments on their surface can be "panned" for transformants carrying antibody activity against a specific antigen.

E. coli surface expression is accomplished by fusing a portion of the lamB protein of E. coli to the polypeptide whose surface expression is desired. Any protein expressed on the cell surface of E. coli can provide the outer membrane spanning signal (surface expression signal) for use in the present invention. More specifically, it has been discovered that amino acid residues 57-181 of mature lamB can act as a signal for surface expression. Such fusion polypeptides are represented by the formula, shown in the direction of amino- to carboxy-terminus:

(F1) $NH_2 - B - Z - COOH$,

wherein B represents the amino acid residue sequence of a polypeptide, preferably heterologous to lamB, and Z represents a sequence of amino acid residues corresponding, and preferably identical, to the sequence of lamB from about residue position 57 to about residue position 181 as shown in Figure 3. The heterologous polypeptide can itself be a fusion protein, and typically contains a periplasmic secretion signal sequence (polypeptide leader), such as the pelB secretion signal, and the like. Thus, a preferred fusion polypeptide is represented by the formula,

(F2) NH_2 - leader - J - Z - COOH

wherein the leader is a sequence of amino acid residues that signal secretion to the periplasm, J is a sequence of amino acid residues of from 6 to 350 residues in length, and Z is as described before in formula (F1). Preferably J is from about 50 to about 150 amino acid residues. More preferably, J is a V_H or V_L as described herein.

Recombinant DNA Molecules

In view of the foregoing, the present invention contemplates. In living organisms, the amino acid residue sequence of a protein or polypeptide is directly related via the genetic code to the deoxyribonucleic acid (DNA) sequence of the structural gene that codes for the protein. Thus, a structural gene for a fusion protein of this invention can be defined in terms of the amino acid residue sequence, i.e., protein or polypeptide, for which it

codes. In addition an important and well known feature of the genetic code is its redundancy. That is, for most of the amino acids used to make proteins, more than one coding nucleotide triplet (codon) can code for or designate a particular amino acid residue. Therefore, a number of different nucleotide sequences may code for a particular amino acid residue sequence. Such nucleotide sequences are considered functionally equivalent since they can result in the production of the same amino acid residue sequence in all organisms. Occasionally, a methylated variant of a purine or pyrimidine may be incorporated into a given nucleotide sequence. However, such methylations do not affect the coding relationship in any way.

Recombinant DNA molecules containing a nucleic acid sequence coding for a fusion polypeptide according to formulas (F1) or (F2) are contemplated by this invention. Thus, the present invention provides for linking a nucleotide sequence coding for any polypeptide immunogen against which antibody production is desired to the outer membrane spanning signal (lamB) polypeptide and/or the secretion signal (pel B) polypeptide as described herein. In preferred embodiments the polypeptide immunogen is a pathogen related immunogen and the conjugate has the capacity to induce the production of antibodies that immunoreact with the pathogen when injected in an effective amount into an animal. Exemplary immunogens of particular importance are derived from bacteria such as B. pertussis, S. typhosa, S. paratyphoid A and B, C. diptheriae, C. tetani, C. botulinum, C. perfringens, B. anthracis, P. pestis, P. multocida, V. cholerae, N. meningitides, N. gonorrhea, H. influenzae, T. palladium, and the like; immunogens derived from viruses such as polio virus, adenovirus,

parainfluenza virus, measles, mumps, respiratory syncytical virus, influenza virus, equine encephalomyeitis virus, hog cholera virus, Newcastle virus, fowl pox virus, rabies virus, pseudorabies virus, feline and canine distemper viruses and the like; rickettsiae immunogen such as epidemic and endemic typhus, and the spotted fever groups, and the like. Immunogens are well known to the prior art in numerous references such as U.S. Patent No. 3,149,036, No. 3,983,228, and No. 4,069,313; in Essential Immunology, 3rd Ed., by Roit, published by Blackwell Scientific Publications; in Fundamentals of Clinical Immunology, by Alexander and Good, published by W.B. Saunders; and in Immunology, by Bellanti, published by W.B. Saunders.

Methods for determining the presence of antibodies to an immunogen in a body sample from an immunized animal are well known in the art.

In preferred embodiments the polypeptide immunogen is a pathogen related immunogen that immunoreacts with, i.e., is immunologically bound by, antibodies induced by the pathogen. More preferably, the pathogen related immunogen is capable of inducing an antibody response that provides protection against infection by the pathogen. Methods for determining the presence of both cross-reactive and protective antibodies are well known in the art.

Expression Vectors

The present invention also contemplates various expression vectors useful in performing, inter alia, the methods of the present invention. Each of the expression vectors is a novel derivative of Lambda Zap.

1. Lambda Zap II

Lambda Zap II is prepared by replacing the Lambda S gene of the vector Lambda Zap with the Lambda S gene from the Lambda gt10 vector, as described in Example 7.

2. Lambda ImmunoZAP H

Lambda ImmunoZAP H is prepared by inserting the synthetic DNA sequences illustrated in Figure 6A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda ImmunoZAP H is described in more detail in Example 8, and its features illustrated in Figures 6A and 7.

3. Modified Lambda ImmunoZAP H

Modified Lambda ImmunoZAP H is prepared by inserting the modified synthetic DNA sequences illustrated in Figure 8A into the above-described Lambda ZAP II vector. The preparation of modified Lambda ImmunoZAP H and the details of the modifications are described in Example 8B. Its features are illustrated in Figure 8A and 8B.

4. Lambda ImmunoZAP L

Lambda ImmunoZAP L is prepared as described in Example 9 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda ImmunoZAP L are illustrated in Figure 9.

Transformants and Vaccines

A host transformed with a recombinant DNA molecule of this invention is also contemplated by this invention. The transformants are useful, not

only in isolating a heterodimeric receptor according to the methods described herein, but also as vaccine strains where the fusion polypeptide immunologically cross-reacts with pathogen-neutralizing antibodies. Methods of formulating and using vaccine strains are described in U.S. Patents No. 4,764,370 and No. 4,337,314.

For heterodimeric molecules that assemble in the cell or in the periplasm, operatively linking the lamB outer membrane spanning signal sequence to the carboxy-terminus of one of the polypeptide chains of the heterodimer, e.g., the heavy chain of a Fab, results in surface expression of the assembled heterodimer.

One of the advantages of the present invention is that a vaccine containing a transformant of this invention can be easily prepared, lyophilized in the presence of appropriate inert, non-toxic carrier(s) (infra) in vials and stored at room temperature without loss of potency. No refrigeration or special storage equipment is required.

The composition of vaccine preparations must be known and consistent. This achieved by using specified amounts of quality-controlled chemical and biological ingredients in their preparation. Methods for the quality control of chemical components are well established in the art and will not be discussed here. Chemical purity in the vaccine preparations is defined as freedom from toxic waste or cellular breakdown products and interfering or spurious immunogenic material. This is assured by working with pure cultures (the vaccine strain free of other cells or virus) and harvesting the cells while the culture is in the logarithmic phase of growth (before the synthesis of autolytic enzymes). The collection and

washing of cells from the medium by physical methods (centrifugation) should leave low molecular weight impurities in the supernatant.

The vaccines of the present invention can be administered to any warm-or cold-blooded animals susceptible to infection with pathogenic microorganisms. Human and non-human animals may benefit as hosts.

Administration can be parenteral, but preferably oral or intranasal, depending upon the natural route of infection. In farm animals, for example, the vaccine may be administered orally, by incorporation of the vaccine in feed or feed water. The dosage administered may be dependent upon the age, health and weight of the recipient, kind of concurrent treatment if any, and nature of the organism. Generally, a dosage of active ingredient will be from about 10^1 to 10^{10} cells per application per host. The preferred dose for intranasal administration would generally be about 10^6 organisms, suspended in 0.05 to 0.1 ml of an immunologically inert carrier. Peroral administration of a vaccine strain of, for example, Salmonella typhi developed according to the method described in this invention would probably require 10^6 to 10^8 organisms suspended in 1-2 mls of, for example, skim milk. The vaccines can be employed in dosage forms such as capsules, liquid solutions, suspensions, or elixirs, for oral administration, or sterile liquid for formulations such as solutions or suspensions for parenteral, intranasal or topical (e.g. wounds or burns) use. An inert, immunologically acceptable carrier is preferably used, such as saline, phosphate buffered saline or skim milk.

Compositions and Kits

Many of the reagents described herein (e.g., nucleic acids such as primers, vectors, and the like) have a number of forms, particularly variably protonated forms, and in equilibrium with each other. As the skilled practitioner will understand, representation herein of one form of a compound or reagent is intended to include all forms thereof that are in equilibrium with each other.

The reagents described herein can be packaged in kit form. As used herein, the term "package" refers to a solid matrix or material customarily utilized in a system and capable of holding within fixed limits one or more of the reagent components for use in a method of the present invention. Such materials include glass and plastic (e.g., polyethylene, polypropylene and polycarbonate) bottles, vials, paper, plastic and plastic-foil laminated envelopes and the like. Thus, for example, a package can be a glass vial used to contain the appropriate quantities of polynucleotide primer(s), vectors, restriction enzyme(s), DNA polymerase, DNA ligase, or a combination thereof. An aliquot of each component sufficient to perform at least one PCR thermocycle will be provided in each container.

Kits useful for producing a template-complement or for amplification of a specific nucleic acid sequence using a primer extension reaction methodology also typically include, in separate containers within the kit, dNTPs where N is adenine, thymine, guanine and cytosine, and other like agents for performing primer extension reactions.

The reagent species of any system described herein can be provided in solution, as a liquid dispersion or as a substantially dry powder, e.g., the plasmids may be provided in lyophilized form.

In one embodiment, the kit is an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L gene primer, and a linking primer defining a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

Another contemplated kit comprises an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, an inside first polypeptide, preferably a V_H, gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a V_H gene. The kit also contains an inside second polypeptide, preferably a V_L, gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of the first polypeptide gene primer, the 3'-terminal priming portion of which comprises a nucleotide sequence homologous to a conserved portion of a second polypeptide gene. The first polypeptide inside and second polypeptide inside primers, when hybridized, form a duplex that codes for a double-stranded DNA molecule containing the before described cistronic bridge for linking the upstream and

downstream cistrons.

Examples

The following examples are intended to illustrate, but not limit, the present invention.

1. Oligonucleotide Primer Design for Producing Dicistronic DNA

A method based on PCR amplification that fuses heavy and light chain sequences has been used to construct a complete antigen binding domain of a Fab protein fragment composed of a heavy and a light chain. Schematic diagrams of an immunoglobulin molecule composed of heavy and light chains containing constant and variable regions is shown in Figure 1. Human heavy chain IgG and human kappa light chain are diagrammatically sketched in Figures 2A and 2B, respectively. To accomplish this procedure, immunoglobulin heavy and light chain primers were designed to produce a region of homology between two polymerase chain reaction (PCR) products. The complementary regions have been shown to hybridize predominantly under conditions where one set of primers ("inside primer pair") is used in a limiting amount relative to the other set of primers ("outside primer pair"). After the 3' ends of the PCR products have hybridized, the DNA polymerase has been shown to extend the ends creating a fusion sequence carrying the unique sequences of both PCR fragments separated by one copy of region X cistronic bridge. A two-step cloning procedure is thus avoided. When the recombinant sequence is then inserted into an expression vector such as ImmunoZAP, a fusion product capable of simultaneously expressing the heavy and light chains can be produced.

The strategy used for producing immunoglobulin heavy and light chain PCR dicistronic DNA is shown schematically in Figure 4. Regions of the immunoglobulin heavy chain coding strand are designated V_H, C_H1, C_H2, and C_H3 corresponding to functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime ('). Regions V_L and C_L are similarly labelled for the kappa light chain. This procedure can also be performed using lambda light chain specific regions. A region, X, unrelated to the natural immunoglobulin sequences, is introduced into the fusion product by attaching X to the 5' ends of both of the C_H1' and V_L inside primers.

Overlapping oligonucleotide primers used in the fusion-PCR reactions to produce dicistronic DNA were designed to encode the following: amino acids of 225 to 230 of the IgG heavy chain hinge region which are common to all human IgG isotypes; an Spe I restriction site; two stop codons; a ribosome binding site; a periplasmic (pelB) leader sequence (Better, et al., Science, 240:1041-1043 (1988); Lei, et al., J. Bacteriol., 169: 4379-4383 (1988)); a Sac I restriction site which encodes amino acids 1 and 2 of the mature kappa light chain; and amino acids 3 to 8 of the mature kappa light chain. The X region was designed to contain a ribosome binding site and a pelB leader to ensure expression of the light chain. Nucleotide sequences for all human and mouse PCR primers, both inside and outside, are listed in Table 7. Primers followed by a prime (') represent non-coding strand sequences.

Table 7

Human and Mouse PCR Primers

	Seq. Id. No.		
		Human	
	(56)	V_H	5'-GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG-3'
	(57)	C_H1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGGT ATTTCATTATGACTGTCTCCTTGCTATTAACTAG TACAAGATTTGGGCTC-3'
	(58)	V_L	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGCT GCCCAACCTGCCATGGCTGAGCTCGTGATGACCC CAGTCTCC-3'
	(59)	C_L'	5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA GCTCTTTGTGACGGGCGAACTC-3'
		Mouse	
	(60)	V_H	5'-AGGTCCAGCTGCTCGAGTCTGG-3'
	(61)	C_H1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGG TATTTCATTATGACTGTCTCCTTGCTATTAACT AGTATACAATCCCTGGGCACAAT-3'
	(62)	V_L	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGC TGCCCAACCTGCCATGGCTGAGCTCGTGATGAC CCAGTCTCC-3'
	(63)	C_L'	5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA-3'

The overlapping regions of the human C_H1' inside and V_L inside primers are illustrated in Figure 5. The heavy chain downstream C_H1' inside primer sequence is written 3' to 5' and the light chain upstream V_L inside primer sequence is written 5' to 3'. The complementary PCR product strands, and not the primer strands, cross-prime to create the dicistronic molecule. Bold nucleotides represent regions where the C_H1' inside primer hybridizes to the 3' end of C_H1 on human IgG heavy chain mRNA or where

the V_L inside primer hybridizes to the 5' end of V_L framework on human kappa light chain cDNA. The amino acid and nucleotides in italics represent changes in sequence from the original pelB leader sequence.

At amino acid 15 of the pelB leader sequence, the codon was changed from CTC to ATC resulting in a conservative amino acid change from a leucine to an isoleucine as shown in Figure 5 and Table 7. Hydrophobic amino acids in the core region of periplasmic leader sequences have been shown to be essential for correct processing of the leader sequence and transport of the mature protein to the periplasm. Oliver, in Neidhardt, R.C. (ed.), Escherichia coli and Salmonella Typhimurium., Am. Soc. Microbiol., 1:56-69 (1987). The nucleotide changes were made to allow for the artifactual insertion of one or two dATPs at the 3' end of the overlapping dicistronic molecules. Thermus aquaticus (Taq) DNA polymerase may add a dATP to the 3' end of the PCR product because of terminal transferase activity. Jiang, et al. Oncogene, 4: 923-928 (1989). The additional dATP would then cause a mismatch between the overlapping PCR products at the 3' terminus and inhibit elongation by Taq DNA polymerase. Sommer, et al. Nucl. Acids Res., 17: 6749 (1989). Therefore, the change to two dTTPs in this position of the oligonucleotide primers would allow proper base pairing if up to two dATPs were added to the 3' terminus of the heavy chain PCR product. The kappa light chain PCR product was designed to terminate at a position where two dTTPs occur 5' of the end of the product and did not require alterations of the nucleotide sequence. Nucleotides were changed in the kappa light chain primer encoding the pelB leader sequence without introducing amino acid changes in

order to decrease the number of mismatches between the primer and the leader sequence of the kappa light chain mRNA as shown in Figure 5 and Table 7.

All primers were synthesized on an Applied Biosystems DNA synthesizer, model 381A, following the manufacturer's instructions.

2. Preparation of a V_H-and V_L-Coding Repertoire

a. Preparation of a V_H-and V_L-Coding Repertoire from a Human cDNA Combinatorial Library

Cloned DNA, previously isolated from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT) was used as a template for preparing a V_H-and V_L-coding repertoire. Mullinax, et al., supra. Briefly, the combinatorial library was prepared by the following approach. Volunteer donors, who had been previously immunized against tetanus but had not received booster injections within the last year, received injections on 2 consecutive days of 0.5 milliliters (ml) of alum-absorbed TT (40 microgram/ml (ug)/ml) (Connaught Laboratories, Swiftwater, Pennsylvania).

One hundred ml of blood was drawn from the volunteers 6 days post injection and anticoagulated with a mixture of 0.14 M citric acid, 0.2 M trisodium citrate, and 0.22 M dextrose. The peripheral blood lymphocytes (PBLs) were recovered and isolated from the whole blood by layering the whole blood on Histopaque-1077 (Sigma, St. Louis, Missouri) and centrifuging at 400 x g for 30 minutes at 25 degrees Celsius (25°C). Isolated PBLs were washed twice with phosphate buffered saline (PBS) (150 mM sodium chloride and 150 mM sodium phosphate, pH 7.2 at 25°C).

Total RNA was then purified from the PBLs (10^6 B cells per ml blood per 100 ml of blood) for an enriched source of B-cell mRNA coding for anti-TT IgG using an RNA isolation kit according to manufacturer's instructions (Stratagene, La Jolla, California) and also described by Chomczynski et al., Anal. Biochem., 162:156-159 (1987). Briefly, the isolated PBLs were homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M beta-mercaptoethanol. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized cells. Ten ml of phenol that had been previously saturated with H_2O was also admixed to the denaturing solution containing the homogenized cells. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuged tube (Fisher Scientific Company, Pittsburgh, Pennsylvania). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described above. Three ml of isopropyl alcohol was added to the re-suspended total cellular RNA and inverted to mix. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA

was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then re-suspended in diethyl pyrocarbonate (DEPC) treated (DEPC-H_2O) H_2O.

Messenger RNA (mRNA) was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY, (1982). Briefly, 500 mg of the total RNA isolated from a PBLs prepared as described above was re-suspended in one ml of 1X sample buffer (1 mM Tris-HCl, (Tris [hydroxylmethyl-aminomethane]) pH 7.5; 0.1 mM EDTA (disodium ethylene diamine tetra-acetic acid), 0.5 M NaCl) and maintained at 65°C for five minutes and then on ice for five more minutes. The mixture was then applied to an oligo-dT (Stratagene) column that was previously prepared by washing the oligo-dT with a solution containing 10 mM Tris-HCl, pH 7.5; 1 mM EDTA, 0.5 M NaCl. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for five minutes at 65°C. The oligo dT column was then washed with 0.4 ml of high salt loading buffer consisting of 10 mM Tris-HCl at pH 7.5, 500 mM sodium chloride, and 1 mM EDTA. The oligo dT column was then washed with 2 ml of 1 X low salt buffer consisting of 10 mM Tris-HCl at pH 7.5, 100 mM sodium chloride, and 1 mM EDTA. The messenger RNA was eluted from the oligo dT column with 0.6 ml of buffer consisting of 10 mM Tris-HCl at pH 7.5, and 1 mM EDTA. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and re-suspended in DEPC H_2O.

The messenger RNA isolated by the above process contains a plurality of different V_H and V_L coding polynucleotides, i.e., greater than about 10^4 different V_H- and V_L-coding genes.

Isolated RNA was converted to cDNA by a primer extension reaction with a first-strand synthesis kit according to manufacturer's instructions (Stratagene) by using an oligo (dT) primer for the light chain and a specific primer, C_H1', for the heavy chain. Mullinax et al., supra. In a typical 50 µl transcription reaction, 5 ug of PBL mRNA in water was first hybridized (annealed) with 200 ng (50.0 pmol) of an oligo (dT) primer for the light chain. In a separate reaction, 5 ug of PBL mRNA in water was first hybridized (annealed with 200 ng (20 pmol) of the heavy chain primer, C_H1', at 65°C for five minutes. Subsequently, the mixture was adjusted to 0.5 mM each of dATP, dCTP, dGTP and dTTP, 50 mM Tris-HCl at pH 8.3, 3 mM $MgCl_2$, 75 mM KCl, 10 mM DTT, 20 units of RNase block II (Stratagene), and 20 units of Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), was added and the solution was maintained for 1 hour at 37°C. PCR amplification of the heavy and light chain sequences was done separately using 0.25-0.5 ug of first-strand synthesis product as template with sets of primer pairs using Taq DNA polymerase as described in Example 3.

The PCR amplified light chain DNA fragments were then digested with Sac I and Xba I and ligated into a modified Lambda Zap II vector as prepared in Example 9 to form a light chain ImmunoZap Library (ImmunoZAP L; Stratacyte, La Jolla, California). The PCR amplified heavy chain DNA was digested with Spe I and Xho I and ligated into a different modified Lambda

Zap II vector as prepared in Example 7 to form a heavy chain ImmunoZap Library (ImmunoZAP H; Stratacyte). The resulting libraries were amplified and the resulting DNA was packaged into bacteriophage with in vitro packaging extract, Gigapack II gold (Stratagene) and used to infect E. coli strain XL1-Blue (Stratagene).

To construct a library for coexpression, the right arm of the heavy chain library phage DNA was digested with Hind III, preserving the left arm of ImmunoZAP H with heavy chain inserts. The left arm of the light chain library phage DNA was digested with Mlu I resulting in a right arm of ImmunoZAP with kappa light chain inserts. Both products were then digested with Eco RI and ligated to create a combinatorial library that encoded human Fab fragments including those specific for TT. Mullinax, et al., supra.

Reactive plaques were first identified by binding to tetanus toxoid as described in Example 11. Bacteriophage from purified reactive plaques were then converted to the plasmid format by in vivo excision with R408 helper phage (Stratagene) following methods described in Example 11 and familiar to one skilled in the art. Short, et al., Nucl. Acids. Res., 16:7583-7600 (1988). The resulting purified plasmid DNA encoding heavy and light chain was then used in PCR reactions as described below in Example 3.

b. Preparation of a V_H- and V_L-Coding Repertoire from mRNA from Tissues and Cells

1) Human

Purified populations of PBLs, other lymphocytes, and hybridomas which express immunoglobulins including IgG, IgM, IgE, IgD, and IgA are used as sources for isolating mRNA encoding

immunoglobulins. PBL's and other immunoglobulin expressing lymphocytes are isolated from either spleen, lymphoid tissue or plasma. Following purification of the cells, total RNA is then purified from these cells using a RNA isolation kit (Stratagene) as described in Example 2a. The purified RNA is then converted to cDNA with a first-strand synthesis kit as described in Example 2a. The resultant cDNA is then used as a template in PCR amplication reactions as described below in Example 3 for the production of dicistronic molecules expressing heavy and light chains.

2) Mouse

Populations of cells described above can be isolated from other mammalian sources such as mouse or rabbit. Both mRNA and rearranged DNA can be isolated as described above and used as templates in PCR amplification reactions. cDNA synthesized from mRNA isolated from a mouse anti-human fibronectin hybridoma (ATCC, CRL-1606) was used as a preferred template for the production of dicistronic molecules expressing heavy and light chain.

c. Preparation of a V_H-Coding Repertoire From Rearranged DNA

Rearranged DNA isolated from PBLs, other lymphocytes, and hybridomas which express immunoglobulins can be used to prepare a V_H-coding repertoire. The amplification procedure for preparing a V_H-coding repertoire using rearranged DNA is performed as described in Example 3.

3. Preparation of DNA Homologs

a. V_H-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 2 from a combinatorial library that encodes human Fab fragments

which bind tetanus toxoid (TT), was used as a template for preparing a V_H-coding double stranded DNA homolog. Human heavy chain, containing both the V_H and C_H1 coding region and designated as Fd, was amplified in a PCR reaction. The amplification was performed in a 100 ul reaction containing 5 nanograms (ng) of the cloned DNA in PCR buffer consisting of the following: 10 mM Tris-HCl, pH 8.3; 50 mM KCl, 1.5 mM $MgCl_2$; 0.001% (w/v) gelatin; 200 mM of each dNTP; 200 nanomolar (nM) of each primer; and 2.5 units of Taq DNA polymerase. The human V_H outside primer and C_H1' inside primer were used as a PCR primer pair for amplification of the heavy chain (Table 7 and Figure 4). The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle (thermocycle) involved denaturation at 94°C for 1.5 minutes, annealing at 54°C for 2.5 minutes and polynucleotide synthesis by primer extension (elongation) at 72°C for 3.0 minutes followed by a return to the denaturation temperature. The resultant amplified V_H-coding DNA homolog containing samples were then gel purified, extracted twice with phenol/chloroform, once with chloroform followed by ethanol precipitation and were stored at -70°C in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA.

To verify the amplification of the heavy chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the heavy chain was approximately 730 base pairs as shown in Figure 10. The V_H-coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with V_L-coding counterparts prepared below for the production of dicistronic DNA molecules having V_H and V_L cistronic portions as illustrated in Example 4.

b. V_L-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 2 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing a V_L-coding double stranded DNA homolog. Human light chain, containing the entire coding region of kappa light chain (V_L and C_L), was amplified using the same PCR conditions described for human heavy chain with the exception that a human V_L inside primer and C_L' outside primer were used as the PCR primer pair (Table 7 and Figure 4). The resultant V_L-coding double stranded DNA homolog was gel purified and stored as described above.

To verify the amplification of the light chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the light chain was approximately 690 base pairs as shown in Figure 10. The V_L-coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with V_H-coding counterparts prepared above for the production of dicistronic DNA molecules as illustrated in Example 4.

4. Preparation of Internally-Primed Duplexes of V_H- and V_L-Coding DNA Homologs

a. Hybridization of V_H- with V_L-Coding DNA Homologs

The V_H- and V_L-coding double stranded DNA homologs prepare in Example 3a and 3b, respectively, were admixed together and denatured at 95°C for 5 minutes to separate the strands of each homolog. The denatured V_H-and V_L-coding DNA strands in the admixture were then annealed at 54°C for 5 minutes to form a V_H- and V_L-coding duplex DNA molecule hybridized at the 3' ends at region X of each original

homolog. One strand of the X region (cistronic) bridge encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon, and a polypeptide leader (pelB) having a translation initiation codon in the same reading frame as the downstream cistron located downstream from the ribosome binding site.

b. Primer Extension to Produce Dicistronic DNA Molecules

The hybridized recombinant V_H- and V_L-coding DNA molecule (internally primed duplex) was subjected to primer extension and then amplified with only the V_H and C_L' primers following the PCR reaction procedure described in Example 3a. This second PCR reaction is schematically represented in Figure 4. The PCR reaction products were gel electrophoresed to verify the presence of the resultant V_H-and V_L-coding dicistronic DNA molecules. The expected size of the dicistronic molecule was about 1390 base pairs and is shown in Figure 10. The resultant V_H- and V_L-coding dicistronic DNA molecules were then ligated into the modified ImmunoZAP H vector (Figure 8) for the construction of expression vectors as described in Example 10.

5. Preparation of Mouse Hybridoma V_H- and V_L-Coding Double Stranded DNA Homologs and Production of Dicistronic DNA Molecules in a Single Amplification Reaction

Mouse hybridoma heavy and light chain cDNA prepared in Example 2b was amplified in a single PCR reaction using the reaction conditions given above with an excess of the outside primers (200 nM concentration of both the mouse V_H primer and C_L'

primer) and a limiting amount of the inside primers (20 nM concentration of both the mouse C_H1' and V_L primer) (Table 7). The resultant mouse heavy and light chain dicistronic molecules were then inserted into a modified ImmunoZAP H for construction of an expression vector as described in Example 10.

6. Preparation of Internally-Primed Duplexes Using a Single Internal Primer that Overlaps Both the V_H and V_L Repertoires

Another approach to producing a library of dicistronic DNA molecules is to use a single internal primer instead of using two separately internal primers. The process of creating a dicistronic molecule comprising an upstream V_H cistron and a downstream V_L cistron is to combine in a PCR buffer the following: a repertoire of V_H genes consisting of at least 10^5 different genes; a repertoire of V_L genes consisting of at least 10^4 different genes; an outside V_H primer; an outside V_L primer; and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The PCR reaction is performed as described in Example 2a.

The 3'-terminal priming portion of a polynucleotide strand (linker) has a nucleotide base sequence homologous to a portion of the primer extension product of one of the outside primers. The 5'-terminal priming portion encodes a nucleotide base sequence homologous to a portion of the primer extension product of the other outside primer. The cistronic bridge coding portion encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon and a polypeptide leader (pelB) having a

translation initiation codon in the same reading frame as the downstream cistron where the initiation codon is located downstream from the ribosome binding site. Polynucleotide strand (linker) primers useful in this invention are listed in Table 8.

Table 8

Polynucleotide Strand (Linker) Primers

Seq. Id. No.			
(64)[1]	1'	5'	GGAGAGTGGGTCATCACGAGCTCAGCCATGGCAGGTTGG GCAGCGATTAATAACAATCCAGCGGCTGCCGTAGGCAAT AGGTATTTCATTATGACTGTCTCCTTGCTATTAACTAGT ACAAGATTTGGGCTC 3'
(65)[2]	2'	5'	GAGCCCAAATCTTGTACTAGTTAATAGCAAGGAGACAGT CATAATGAAATACCTATTGCCTACGGCAGCCGCTGGATT GTTATTAATCGCTGCCCAACCTGCCATGGCTGAGCTCGT GATGACCCACTCTCC 3'

[1] Primes mRNA (sense strand) of heavy chain C_H1 region; antisense strand of light chain V_L with dicistronic bridge in between heavy and light chains will be in the same relative orientation as given in the example.

[2] Primes antisense strand of heavy chain C_H1 regions; and sense strand of light chain V_L region with dicistronic in between heavy and light chains will be in the same relative orientation as given in the example.

The resultant single step internally primed dicistronic DNA molecule can then be ligated into modified ImmunoZAP H for construction of an expression vector as described in Example 10.

7. Preparation of Lambda Zap II Expression Vector

The vector Lambda Zap™ II (Stratagene) is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluescript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, 1988, by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard protocols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, NY, 1987.

8. Preparation of V_H-Expression Vectors, ImmunoZAP H and Modified ImmunoZAP H, Construction

a. ImmunoZAP H

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second,

it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. The alternative, plasmid cloning vectors, are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of V_H-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the V_H-coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the V_H-coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6A. The individual single-stranded polynucleotides (N_1-N_{12}) are shown in Table 9 below.

Table 9

Seq. Id. No.		
(66)	N1)	5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
(67)	N2)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(68)	N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'

(69) N4) 5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
(70) N5) 5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(71) N6) 5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'
(72) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(73) N8) 5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3'
(74) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(75) N11) 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(76) N12) 5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(77) N10-5) 5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

Polynucleotides 2, 3, 9-4', 11, 10-5', 6, 7 and 8 were kinased by adding 1 μl of each polynucleotide (0.1 ug/μl) and 20 units of T_4 polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 10 mM beta mercaptoethanol, 500 ug/ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides, 20 ng, of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 6A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 μl of the above reaction to a solution containing 50 mM Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining

the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 μl of the above reaction, 4 μl of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a lambda Zap II vector prepared in Example 7 that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract (Stratagene). The packaged ligation mixture was plated on XL1-blue cells (Stratagene). Individual Lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer (Stratagene). This in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described in by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacture's instructions in the AMV Reverse Transcriptase ^{35}S-ATP sequencing kit (Stratagene). The sequence of the resulting V_H expression vector is shown in Figure 6A and Figure 7.

b. Modified ImmunoZAP H

To create a fusion-PCR library from hybridoma RNA for expressing the plurality of V_H-coding DNA homologs in an E. coli host cell, a vector based on the ImmunoZAP H vector described above was constructed. The procedure for constructing the

vector was performed as described above with the following modifications: elimination of the Sac I site between the T_3 polymerase and Not I sites and changing the nucleotide base residue sequence from AAA to CAG which resulted in an amino acid residue change from lysine to glutamine as shown in Figure 8A and 8B.

The individual single-stranded polynucleotides (N_1, N_4, N_6 and N_7), which were modified from their counterparts listed in Table 9, are listed in Table 10 below.

Table 10

Seq. Id. No.		
(78)	N1)	5' AGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
(67)	N2)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(68)	N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'
(79)	N4)	5' AGGTGCAGCTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
(70)	N5)	5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(80)	N6)	5' CAGCTGCACCTGGGCCATGGCTGGTTGGG 3'
(72)	N7)	5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(81)	N8)	5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGCGGCCGC 3'
(74)	N9-4)	5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(75)	N11)	5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(76)	N12)	5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(77)	N10-5)	5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

The modified ImmunoZAP H vector was created to eliminate an unnecessary Sac I site in the ImmunoZAP H vector, (Example 9), when the heavy and light chain vectors were combined. The modifications also improved the efficiency of secretion of positively changed amino acids in the amino terminus of the expressed protein. Inouye et al., Proc. Natl. Acad. Sci. USA, 85:7685-7689 (1988).

9. Preparation of V_L Expression Vector ImmunoZAP L Construction

To express the plurality of V_L coding polynucleotides in an E. coli host cell, a vector was constructed that placed the V_L coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the V_L polynucleotide. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6B. The individual single-stranded polynucleotides (N_1-N_8) are shown in Table 9 above.

Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1 μl of each polynucleotide and 20 units of T_4 polynucleotide kinase to a solution containing 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DDT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual

polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 µl of the above reaction to a solution containing 50 µl Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 µl of the above reaction, 4 µl of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a Lambda Zap II vector prepared in Example 7 that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract and the packaged ligation mixture was plated on XL1-Blue cells as described in Example 8a. Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol as described in Example 8a. This in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a phagemid vector to allow easy manipulation and sequencing and also produces the phagemid version of the V_L expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase ^{35}S-dATP sequencing kit (Stratagene). The sequence of the

PCT/US92/01475

resulting V_L expression vector is shown in Figure 6B and Figure 9.

The V_L expression vector used to construct the V_L library was the phagemid produced to allow the DNA of the V_L expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap V_L expression vector (Figure 9).

10. Construction of V_{HL} Expression Vectors and Library

a. Ligation of Dicistronic DNA Molecules with Modified ImmunoZAP H

In preparation for cloning a library enriched in V_H-V_L-coding (V_{HL}) dicistronic DNA molecules, PCR amplified products (human or mouse) prepared in Examples 4, 5 and 6 (50 mM NaCl, 25 mM Tris-HCl, pH 7.7, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 100 ug/ml BSA, at 37°C were digested with restriction enzymes Xho I and Xba I at a concentration of 60 units of enzyme per ug of DNA, and purified on a 1% agarose gel. After gel electrophoresis of the digested PCR amplified dicistronic DNA molecules, the region of the gel containing the DNA fragments of approximately 1360 base pairs in size was excised, purified using Gene-Clean (BIO 101, La Jolla, California), ethanol precipitated and resuspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 4°C to 1 ug of modified ImmunoZAP H vector, prepared in Example 8b, (Stratagene) previously digested with Xho I and Xba I. A portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold

packaging extract (Stratagene) and the packaged material was plated on a permissive E. coli (strain XL1-blue) lawn to generate plaques. The library was determined to consist of predominantly V_{HL} with less than 5% non-recombinant background.

b. Screening of Antibody-Producing Plaques

1) Human

To screen for expression of V_{HL} dicistronic molecules, E. coli were infected to yield approximately 100 plaques per plate. Replica filter lifts of the plaques on an agar plate were produced by overlaying a nitrocellulose filter that had been soaked in 10 mM isopropyl beta-dithiogalactopyranoside on each plate with transfer for 15 hours at 23°C. For detection of V_{HL} antibody fragment expression, the filters were screened with rabbit anti-human heavy and light chain antibodies followed by goat anti-rabbit antibody coupled to alkaline phosphatase (Cappel Laboratories, Malvern, Pennsylvania). The detection of immunoreactive product confirmed the presence and expression of V_{HL} antibody fragments.

To identify human DNA clones expressing antibody that bound TT, plaques were plated and proteins expressed as described above. Replica filters were incubated with 0.2 nM ^{125}I-tetanus toxoid and washed. Positive plaques were identified by autoradiography and isolated. The frequency of positive clones in the library was equivalent to (number of positive clones)/[number of plaques screened) X (fraction of plaques expressing V_{HL}). Concentrated nonadsorbed tetanus toxoid was iodinated with sodium iodide ^{125}I (ICN, Irvine, California) by the Chloramine-T method as described in Botton et al., Biochem. I., 133:529-539 (1973) and available in a kit (Iodo-Beads, Pierce, Rockford, Illinois).

Human DNA clones were re-plated at approximately 100 phage per plaque side by side with the parental phage that were used as templates for PCR amplification and screened in the primary antigen binding screen. The results of the screening procedure are seen in Figure 11. Similar signals between the parental clones and the V_{HL} dicistronic DNA molecules demonstrated that the sequence differences introduced with the C_H1' and V_L primers did not adversely affect gene expression. Also, it should be noted in Figure 11 that a random parental clone that did not react with tetanus toxoid, 7G1, was unreactive before and after the PCR dicistronic fusion, as was the control ImmunoZAP H vector (IZ H).

2) Mouse

Mouse antibody-producing plaques prepared in Example 7 were screened for antibody expression with rabbit anti-mouse heavy and light chain antibody (Cappel Laboratories) as described above.

11. Characterization of Cloned Dicistronic V_{HL} Repertoire in Expression Library

a. Verification of Presence and Size of Cloned Dicistronic V_{HL} Repertoire

Bacteriophage from purified reactive plaques prepared in Example 10b were converted to the plasmid format by in vivo excision with R408 helper phage according to manufacturer's protocol (Stratagene) and also described in Short et al., Nucl. Acids Res., 16:7583-7600 (1988). In the in vivo excision protocol, the cloned insert from the ImmunoZAP H vector was converted into a phagemid vector to allow easy manipulation and sequencing. Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing

500 ul of a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% (w/v) gelatin and 20 ul of chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells (A_{600} = 1.00) and 1 ul of R408 helper phage (1 x 10^{10} plaque forming units (pfu)/ml) were incubated at 37°C for 15 minutes. After a 4 hour incubation in Luria-Bertani (LB) broth and heating at 70°C for 20 minutes to heat kill thc XL1-blue cells, the phagemids were re-infected into XL1-Blue cells and plated onto LB plates containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al., Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with Xho 1 and Xba 1. The detection of 1390 base pair fragment on an agarose gel confirmed the presence of a V_{HL} dicistronic molecule insert.

b. Sequencing of Plasmids from Expression Library

Clones containing the putative V_{HL} insert were sequenced using reverse transcriptase according to the general method described by Sanger et al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instructions in the AMV reverse transcriptase ^{35}S-dATP sequencing kit (Stratagene).

Nucleotide sequence analysis of several fusion clones indicated that the sequence of the fusion region was identical to that shown in Figure 5, proving that the clones were actually generated through a fusion PCR intermediate.

c. Advantages of Fusion-PCR to Produce Dicistronic DNA Molecules

PCR amplification can, therefore, be used to fuse sequences responsible for encoding subunits of a heterodimeric protein together into a single DNA fragment that can then direct the expression of both subunits from one expression vector. In the case of antibodies, if the source of nucleic acid template comes from hybridoma mRNA, there is only one heavy and light chain sequence to choose from, and thus the heavy:light pair is a "natural" pair.

However, if spleen, peripheral blood B-cell, or other lymphocyte mRNA is used as the source of template, the PCR fusion reaction to form a dicistronic DNA molecule can randomly pair heavy and light chains from different cells, producing a combinatorial library. In such a library, only a small fraction of the clones contain the original heavy and light chain pairs. This may not be a problem if the desired natural pair is well represented in the original B-cell population, as is the case with hyperimmunized donors. However, if one wishes to find a naturally occurring rare specificity in a combinatorial library, one may have to screen an large number of clones.

The fusion method presented here may offer a solution to the random combinatorial problem. If one begins with a very dilute population of B-cells (possibly in a medium that limits diffusion), it may be possible for the dicistronic event to occur between naturally paired heavy and light chain sequences before significant mixing between B-cell RNA occurs. Thus, the fused heavy and light chain sequences would be the original pairs, and the resulting library would express predominantly the naturally occurring antibody specificities. Such a library would be highly preferable when rare natural specificities are sought.

Another advantage to this method is that only one vector and one cloning step are necessary. This saves a substantial amount of time, resources, and effort. Moreover, the ease of the single PCR reaction greatly simplified the process of going from B-cell RNA to an E. coli library, making this approach a noteworthy alternative to standard hybridoma technology.

12. Production of An Expression Vector for Fusing the LamB Outer Membrane Spanning Signal to a Polypeptide

The following PCR primers are used to produce a DNA segment encoding the surface expression signal amino acid residue sequence of lamB, (i.e., residue positions 51-184 as shown in Figure 3):

Table 11
LamB Primers

Seq. Id. No.		
(82)	upstream[1]	5' TTACTAACTAGTTTCTATTTCGACACTAACGTG3'
(83)	downstream[2]	5' TTAGATCTAGATTTCCATCTGCGCTAAACGCAC3'

[1] Underlined sequence designates the location of an Spe I restriction site.

[2] Underlined sequence designates the location of an Xba I restriction site.

The primers are mixed pairwise with genomic DNA used from E. coli having the lamB gene as template. The amplified DNA segment is purified by preparative agarose gel electrophoresis, digested with Spe I and Xba I restriction endonucleases, and

subsequently ligated into ImmunoZap H/L at the Spe I restriction sites flanking the decapeptide tag sequence, i.e., the decapeptide tag sequence is replaced with the outer membrane spanning signal sequence.

The immunoZAP vector (H/L) is created from the heavy and light chain libraries, prepared in Examples 8 and 9, respectively, by fusing the vectors at the Eco R1 site as follows. DNA is purified from the light chain library and restriction digested with Mlu 1 and Eco R1. This cleaves the DNA from the left arm of the vector into several pieces while leaving the right arm with the light chain inserts intact. DNA is purified from the heavy chain libraries and restriction digested with Hind III and Eco R1. This cleaves the DNA from the right arm of the vector into several pieces while leaving the left arm with the heavy chain inserts intact. The intact left arm of the heavy chain vector containing the heavy chain inserts and right arm of the light chain vector containing the light chain inserts are then mixed and ligated at the common Eco R1 restriction site. The resultant ImmunoZAP H/L vector is shown in Figure 12. The ligations and packaging are as described in Example 2 to create the ImmunoZAP H/L library.

A DNA segment coding for a preselected polypeptide, such as a V_H, can then be ligated into the lamB-modified ImmunoZap H expression vector at position between, and is the same reading frame with, the pelB leader and the lamB signal sequences. The vector thus produced expresses the preselected polypeptide as a double-fusion protein, i.e., having pelB leader and lamB surface expression signal polypeptide segments operatively linked to the preselected polypeptides amino- and carboxy-termini,

respectively.

To increase the distance of the preselected polypeptide from the surface membrane of E. coli which results in decreased steric hinderance and competition of the preselected polypeptide with the lipopolysaccharide coat of E. coli, inserts of various lengths can be constructed in the recombinant plasmid vector (Tables 12 and 13).

The amino acid residue sequence of the inserts A, B, C or D following expression in E. coli as described in Example 10 are listed in Table 12.

Table 12

Insert Amino Acid Residue Sequence

Seq. Id No.	
(84)[1]	GluProLysSerCysAspLysThrHisThrSerProProAla ProAlaProGluLeuLeuLysSerSerPheTyrPheAspThr
(85)[2]	ProLysSerCysAspLysThrHisThrGluProLysSerThr AspLysThrHisThrSerProProAlaProAlaProGluLeu LeuLysSerSerPheTyr
(86)[3]	ProLysSerCysAspLysThrHisThrSerLysSerSerPhe TyrPheAsp
(87)[4]	GluProLysSerCysAspLysThrHisThrSerTyrPheTyr AspValProAspTyrGlySerLysSerSerPheTyrPheAsp Thr

[1] Insert A: Moves Spe 1 site, retains native IgG1 upper hinge region; retains original lamB sequence.

[2] Insert B: Moves Spe 1 site, retains original IgG1 upper hinge region, retains original lamB sequence.

[3] Insert C: Moves Spe 1 site, retains original IgG1 upper hinge region, duplicates 10 amino acids in the upper hinge region.

[4] Insert D: Moves Spe 1 site, retains native IgG1

upper hinge region; retains original lamB sequence.

Table 13

Insert Primers

Seq Id No.		
(88)[1]	5'	GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG 3'
(89)[2]	5'	GTCCACCGGCCCCAGCACCTGAACTCCTGAAGAGCAGC TTCTAT 3'
(90)[3]	3'	CTCGGGTTTAGAACACTGTTTTGAGTGTGATCAGGTGGC CGGGGTCGTGGAC 5'
(89)[4]	5'	GTCCACCGGCCCCAGCACCTGAACTCCTGAAGAGCAGC TTCTAT 3'
(90)[5]	3'	CTCGGGTTTAGAACACTGTTTTGAGTGTGATCAGGTGGC CGGGGTCGTGGAC 5'
(91)[6]	5'	GTGACAAAACTCACACTAGTAAGAGCAGCTTCTAT 3'
(92)[7]	3'	CTCGGGTTTAGAACACTGTTTTGAGTGTGATCA 5'
(93)[8]	5'	GTGACAAAACTCACACTAGTTACCCGTACGACGTTCCGGAC TACGGTTCTAAGAGCAGCTTGTAT 3'
(94)[9]	3'	CTCGGGTTTAGAACACTGTTTTGAGTGTGATCATCA 5'
(95)[10]	3'	CACGCAAATCGCGTCTACCTTAAGATCGTTCCTCTGTCA GTATTACTTTATGGATAACGGATGCCGTCGGCGACCTAA CAATAA 5'
(96)[11]	5'	GCCTACGGCAGCCGCTGGATTGTTATTAATCGCTGCCCA ACCTGCCATGGCTGAGCTCGTGATGACCCATGCTCC 3'
(97)[12]	5'	TCCTTCTAGATTACTAACACTCTCCCCTGTTGAAGCTCT TTGTGACGGGCGAACTC 3'

[1] 5' heavy chain V_H primer used for all constructions.

[2] 5' lamB overlapping primer for insert A.

[3] 3' heavy chain C_H1 overlapping primer for insert A.

[4] 5' lamB overlapping primer for insert B.

[5] 3' heavy chain C_H1 overlapping primer for insert B.

[6] 5' lamB overlapping primer for insert C.

7 3' heavy chain C_H1 overlapping primer for insert C.
8 5' lamB overlapping primer for insert D.
9 3' heavy chain C_H1 overlapping primer for insert D.
10 3' lamB overlapping primer with 5' light chain primer.
11 5' light chain overlapping primer with 3' light chain prime.
12 3' lamB overlapping primer with 5' light chain primer.

The inserts between the heavy chain and lamB sequences are made using the PCR-fusion procedure for producing dicistronic DNA as prepared in Examples 2 and 3 with the following exceptions. The light chain and lamB sequences are fused together using the outside primers and limiting amounts of the inside primers (Table 13). The resultant PCR products are gel purified using Gene Clean (BIO 101) as described in Example 10 before PCR-fusing it to the heavy chain using only outside primers (Table 13). The resultant PCR-fusion product consists of V_H-insert A, B, C or D-lamB-light chain. The region inserted by the PCR primers between the lamB and light chain creates the same dicistronic bridge previously inserted between the heavy and light chain DNAs. This product is ligated with the modified ImmunoZAP H vector restriction digested with the enzymes Xho I and Xba I as prepared in Example 10. After insertion, the dicistronic message encoded by the DNA allows expression of the heavy chain and lamB as a fusion protein and the light chain as a separate protein.

Surface expression is accomplished by transforming the recombinant plasmid vector into an E. coli strain, lacking its endogeneous lamB gene, thus avoiding competition between the recombinant lamB

signal-containing product and a normal lamB gene product for available membrane-spanning sites. A preferred host is a lamB deletion mutant of the E. coli SURE strain. The E. coli SURE strain is commercially available from Stratagene.

13. Production of Antigen-Specific B Cells

a. In Vitro Immunization

1) Preparation of T Cell Replacing Factor, s-PWM-T

Blood was collected from healthy donors and PBLs were isolated as described in Example 2. Isolated PBLs were then fractionated into T and non-T cells by AET-SRBC (2-aminoethylthiouronium bromide-sheep red blood cell) rosetting according to the procedure described by Callard. Callard et al., Eur. J. Immunol., 11, 206 (1981). Briefly, the isolated PBLs were treated with a 1% suspension of AET-modified sheep red blood cells. The rosette was purified over a Ficoll gradient and the red blood cells removed by hypotonic lysis.

The procedure for preparing the T cell replacing factor, s-PWM-T, was performed as described by Danielson. Danielson et al., Immunol., 61:51-55 (1987). The resultant enriched T cell population was suspended in RPMI-1640 medium (Gibco Laboratories, Santa Clara, California) supplemented with 1% (v/v) non-essential amino acids, 4 mM L-glutamine, streptomycin (50 ug/ml), penicillin (50 IU/ml) and 10% heat-inactivated human AB serum at a concentration of 2×10^6 cells/ml, and irradiated (2000 rads; 1 rad = 0.01 Gy). Following irradiation, the T cells were activated by treatment with 10 ug of pokeweed mitogen (PWM)/ml (Sigma) for 24 hours at 37°C. The supernatant was collected and stored at 4°C. PWM

activation of T cells results in secretion of gamma interferon, interleukin-2 (IL-2) and various undefined B cell growth factors into the medium. Growth factor containing supernatant from the PWM-treated T cells, hereinafter designated s-PWM-T, was collected and added to lymphocyte cell cultures prepared below.

2) Preparation of In Vitro Immunization Cultures

The procedure for in vitro immunization of PBLs was performed as described by Borrebaeck. Borrebaeck et al., Proc. Natl. Acad. Sci. USA, 85:3995-3999 (1988). Human PBLs, isolated as described in Example 2, were resuspended to a concentration of 1 X 10^7 cells/ml in serum-free, supplemented RPMI-1640, prepared as described above, containing 2.5 mM L-leucine methyl ester hydrochloride (Leu-OMe) from a 0.5 M stock solution prepared in water. (Sigma Chemical Co., St. Louis, Missouri). The cells were incubated at room temperature for 40 minutes and then washed three times in RPMI-1640 containing 2% heat-inactivated human serum. Cell recovery after treatment with Leu-OMe ranged from 30-90%. The treatment with Leu-OMe was performed to effect the removal of a Leu-OMe-sensitive subpopulation leaving a population of cells that respond to T-cell dependent antigen stimulation in vitro.

Leu-OMe-treated PBLs were immunized in vitro with either keyhole limpet hemocyanin (KLH) (Sigma) or tetanus toxoid (TT) (Example 2). For this protocol, and for the subsequent ELISPOT assays, the Leu-OMe-treated T cells were first suspended in supplemented RPMI-1640, containing 50 uM beta-mercaptoethanol, 10% heat-inactivated human AB serum, 30% (v/v) s-PWM-T, and antigen (1-1000ng/ml). For determination of total

immunoglobulin content in the culture supernatants, the cells were maintained in heat-inactivated fetal bovine serum instead of human AB serum. The antigen-treated Leu-OMe-treated PBLs were then plated at a concentration of 2 X 10^6 cells/ml in a 4-ml (six-well plates) or 30-ml (75-cm^2 flask) and maintained at 37°C in 5% CO_2 for three days. The cells were pelleted and washed one time with RPMI-1640 supplemented medium prepared above lacking antigen to effect the removal of antigen. The washed antigen-treated cells were resuspended in fresh medium containing s-PWM-T, but lacking antigen. The cells were thereafter cultured for three to four more days for a total maintenance period of six to seven days, at which time the levels of antigen-specific antibody and/or the number of antigen-specific antibody secreting cells were determined by ELISA and ELISPOT assays, respectively.

b. Immunoassays

1) ELISA Assay for Determining the Levels of Antigen-Specific Antibody

The antigen-specific immunoglobulin (IgM and IgG) secreted into the medium from antigen-treated PBLs prepared above was determined by ELISA. Briefly, 100 ng/ml of antigen, either KLH or TT, diluted in PBS, pH 7.5, was added to individual wells of 96-well microtitre plates. The plates were allowed to stand at room temperature for 16 to 18 hours. After removing unabsorbed proteins, wells were blocked with 0.2% gelatin-PBS for 1 hour at room temperature.

One hundred ul of culture medium samples were added to the antigen-coated wells and incubated at room temperature for 1 hour. The wells were then rinsed three times with PBS-containing 0.05% Tween 20. Alkaline phosphatase (AP) conjugated to isotype-specific antibodies (goat-anti-human IgM or IgG) (1

ug/ml-final concentration) (Boehringer Mannheim, Indianapolis, Indiana) was diluted in 50 mM PBS, pH 7.5, containing 1.5 M sodium chloride and 0.1% Tween 20 and 100 ul of diluted AP antiglobulin conjugate were then added to each well and maintained at room temperature for one hour or at 4°C overnight. The wells were then rinsed three times with PBS 0.05% Tween 20.

The washed wells were then inverted to remove the remaining buffer. A 1 mg/ml solution of PNPP (Sigma) p-nitro phenylphosphate dissolved in PBS was then added to each well for detection of antigen-specific antibodies and optical density measured at 405 nm.

2) ELISPOT Assay for Determining the Number of Antigen-Specific Antibody-Secreting Cells

ELISPOT assays are performed as described by Czerkinsky. Czerkinsky et al., J. Immunol. Methods, 65:190-121 (1983). For measuring the number of antigen-specific antibody-secreting cells in the in vitro immunized PBL cultures ELISPOT was performed. For this assay, 3.5 centimeter diameter polystyrene petri dishes (Falcon, Oxnard, California) were filled with 1.5 ml of either KLH or TT antigen at a concentration of 1 ug/ml. Borrebaeck et al., supra. The plates were washed as described for the ELISA assay. The antigen-coated plates were then blocked with 0.2% gelatin at 37°C. Lymphocytes (10^5 to 10^6) were added to each dish and allowed to incubate undisturbed overnight at 37°C. The cells were removed and the plates were washed twice with cold PBS and then maintained for 10 minutes with cold 10 mM EDTA-PBS. The plates were then rinsed three times with PBS containing 0.5% Tween-20. Antigen-specific human

antibody was detected with alkaline phosphatase-goat anti-human IgG or IgM, followed by the addition of enzyme substrates 150 ug/ml BCIP (5-bromo-4-chloro-3-indolyl phosphate), and 300 ug/ml NBT (nitroblue tetrazolium), dissolved in molten 1% agarose in PBS. The plates were then incubated for one to several hours at room temperature, after which time, blue spots appeared corresponding to the positions of antigen-specific antibody secreting cells.

The frequency of antigen-specific B cells was determined as (number of antigen-specific antibody secreting cells)/(number of B cells added per plate). In vitro immunization was demonstrated by an increase in the frequency of antigen-specific B cells, in response to antigen stimulation. The number of B cells added to each ELISPOT plate was assumed to be approximately 10% of the total number of Leu-OMe-treated PBLs based on immunofluorescence analysis of Ohlin. Ohlin et al., Immunolology, 66:485-90 (1989). The total number of lymphocytes was determined by trypan exclusion. The results of these assays are described below.

3) Panning to Increase the Frequency of Antigen-Specific B Cells

PBLs, prepared in Example 2, were treated with Leu-OMe and resuspended in supplemented RPMI-1640 medium containing 2% fetal bovine serum as described above. Approximately 1 to 10 x 10^6 Leu-OMe-treated PBLs were added to polystyrene petri dishes, previously coated with 1 μg/ml of either KLH or TT antigen and blocked with 0.2% gelatin. The cells were then maintained at 4°C for 1 hour. After the non-adherant cells were decanted, the plates were washed three times with chilled medium and the non-adherent fractions were pooled.

Depletion of antigen-specific B cells was demonstrated by culturing the non-adherent cells in the presence of s-PWM-T, as described above, for 6 days. The number of antigen-specific antibody producing cells was then determined by the ELISPOT assay. The number of B cells which adhere under the conditions described above was determined using two different methods. An enriched population of B cells was obtained by rosetting with AET-treated sheep red blood cells. The non-rosetting cells were then panned on autologous plasma-coated petri dishes, and the non-adherent lymphocytes (B cells) recovered. In one set of experiments, the B cells were labelled overnight with ^{35}S-methionine, panned as described above, and the percent radioactivity adhering to the dishes determined. In the second set of experiments, the number of purified cells which adhered was determined microscopically using an ocular grid. The results of the experiments are described below.

4) Panning In Vitro Immunized Cells

Panning and in vitro immunization were combined to enrich the frequency of antigen-specific B cells beyond the level which can be achieved with either technique alone. In vitro stimulated lymphocytes were cultured as described above for 5 days, transferred to fresh medium and panned as described on antigen coated dishes. The results of these experiments are described below.

c. Results of In Vitro Immunization and Cell Panning

Using KLH and tetanus toxoid (TT) as model antigens in the above-described procedure resulted in a 2-3 fold increase in the frequency of both TT- and KLH-specific B cells. The frequency of KLH-specific B cells was considerably influenced by

resuspending the cells in fresh media, containing T cell supernatant and lacking antigen, several days after exposure to antigen. As Table 14 shows, a 7 to 17-fold increase in the number of antigen-specific B cells was observed when the cells were pulsed with antigen for 2 to 3 days. Exposing the cells for the entire culture period (6-7 days), on the other hand, resulted in an average increase of only 3-fold.

Table 14

Expansion of KLH-specific B cells*

Expt.	Donor	Day Media Exchange/KLH Removed						
		1	2	3	4	5	6	7
1	1	1	17	4	2	16	6	nd
2	1	nd	>9	1	3	4	4	nd
	2	nd	1	7	1	2	1	nd
3	2	nd	14	2	nd	1	nd	2
	3	nd	>15	>11	nd	1	nd	4

* ratio = (# anti-KLH Ig secreting cells cultured with s-PWM-T and KLH)/(# anti-KLH Ig secreting cells cultured with s-PWM-T) detected with ELISPOT

As expected for induction of a primary immune response, the KLH-specific antibodies secreted were of the IgM isotype. Antigen pulsing resulted in an average increase of 9-fold and a mean frequency of 3.3 x 10^{-3} (Table 15). These results indicate that primary immunization of naive human B cells can give rise to frequencies of antigen-specific B cells which are comparable to those found when B cells were collected from human donors boosted with TT. Mullinax

et al., supra.

Table 15
Frequency of KLH-specific Cells
After Primary Immunization

# Expts	# Donors	Antigen Pulse [Ag] (ng/ml)	Pulse Time (days)	Freq anti-KLH IgM B Cells Background (range)	+ Ag (range)
4	6	400-800	2-3	$3.8X10^{-4}$ ($0.6-7.9X10^{-4}$)	$3.3X10^{-3}$ ($0.9-8.7X10^{-3}$)

In addition to in vitro immunization, cell panning techniques have also been developed for the enrichment of antigen-specific B cells. Table 16 summarizes the degree of enrichment observed for a single cycle of panning against a model antigen.

Table 16
Enrichment of Antigen-Specific B Cells

Expt. #	% Anti-TT Ig Secreting Cells Adhered	Expt. #	% B Cells Adhered	Enrichment Factor
1	100% ± 0%	3	1.5% ± 0.6%	67
2	90% ± 14%	4	10% ± 1.9%	9

Peripheral blood lymphocytes from unboosted donors were panned on TT- and gelatin-coated petri dishes and the number of TT-specific B cells in the non-adherent cell population determined. In experiments 1 and 2, 100% and 90% of the anti-TT antibody secreting cells, respectively, were depleted when panned on TT plates, while only 28% and 8% were depleted when panned on gelatin (not shown). The number of purified B cells which adhere under analogous

conditions ranged from 1.5 to 10% and was determined either by labeling the cells with ^{35}S-methionine (expt. 3) or by examining the adherent cells microscopically with an ocular grid (expt. 4). These preliminary results indicate that a single cycle of cell panning can be used to increase the frequency of antigen-specific B cells by at least 9-fold, and possibly as high as 67-fold. It should be possible to further deplete B cells which bind non-specifically or with low affinity to antigen by performing sequential isolations or by altering the epitope density of the solid matrix.

By combining the results found in Tables 15 and 16, it is evident that cell panning can be used alone or in combination with in vitro immunization to increase the frequency of antigen-specific B cells in the naive repertoire by 2 to 3 orders of magnitude. Analysis of the non-adherent cells recovered after panning, before and up to 7 days after culturing with T cell supernatant (s-PWM-T), indicates that the majority of KLH-specific antibody producing cells are depleted when panned at 0 to 5 days. As Table 17 indicates, panning at days 6 and 7 (peak of antibody production) is inefficient, possibly due to either down-modulation of surface IgM receptors or interference by secreted anti-KLH antibody. To recover the greatest enrichment antigen-specific B cells, panning should be performed at day 5 to ensure maximal clonal expansion.

Table 17

# of Anti-KLH IgM-Secreting Cells in the Non-Adherent Fraction

	Panning Antigen	
Day Panned	KLH	Gelatin
2	1	9

4	1	17
5	2	17
6	17	20
7	18	5

These studies have demonstrated, with model antigens, that in vitro immunization or cell panning can be used to increase the frequency of antigen-specific B cells by at least 10-fold. Preliminary results indicate that the two techniques can be combined to give rise to frequencies which are comparable to those of the lymphocyte population used to construct the TT-specific library (10^{-3}). These techniques may obviate the requirement for in vivo immunization, thereby eliminating one of the major obstacles to the routine production of human monoclonal antibodies. By cloning human immunoglobulin sequences from E. coli expression libraries, the difficulties encountered in immortalizing antibody producing cell lines are avoided as well. Thus, preparing immunoexpression libraries from enriched populations of naive B cells should render it possible to generate human monoclonal antibodies against a variety of antigens of therapeutic and diagnostic interest.

The foregoing is intended as illustrative of the present invention but not limiting. Numerous variations and modifications can be effected without departing from the true spirit and scope of the invention.

What is claimed is:

1. A method of producing a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, said upstream and downstream cistrons encoding respective first and second polypeptides of a heterodimeric receptor, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, a repertoire of first polypeptide genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer, said inside first gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a first gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, a repertoire of second polypeptide genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer, said inside gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said first gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a second polypeptide-coding gene;

said first inside and second inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding (i) at least one stop codon in the same reading frame as said upstream cistron, (ii) a ribosome binding site

downstream from said stop codon, and (iii) a polypeptide leader having a translation initiation codon in the same reading frame as said downstream cistron, said initiation codon located downstream from said ribosome binding site;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form;

(e) separating the double stranded DNA homologs produced in steps (b) and (d);

(f) hybridizing the separated strands to form a plurality of internally-primed duplexes; and

(g) subjecting the internally-primed duplexes to conditions for primer extension to produce a plurality of different dicistronic DNA molecules, each containing a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said first polypeptide- or second polypeptide-coding DNA homologs, and said downstream cistron comprising the other of said first polypeptide- or second polypeptide-coding DNA homologs.

2. The method of claim 1 wherein steps (a)-(d) are performed concurrently in one reaction vessel.

3. The method of claim 1 wherein said signals for the initiation of translation of the downstream cistron are located downstream from the stop codon and include a ribosome binding site and a translation initiation codon encoding the first amino acid residue of a polypeptide leader, said codon located in the same reading frame as the downstream cistron.

4. A method of producing dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, said upstream and downstream cistrons encoding respective first and second polypeptides of a

heterodimeric protein, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, first polypeptide-encoding genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer, said inside first gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of said first polypeptide gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, second polypeptide-encoding genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer, said inside gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said first gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of said second polypeptide-coding genes;

said first inside and second inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding (i) at least one stop codon in the same reading frame as said upstream cistron, and (ii) signals for the initiation of translation of the downstream cistron;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form;

(e) separating the double stranded DNA homologs produced in steps (b) and (d);

(f) hybridizing the separated strands to form a plurality of internally-primed duplexes; and

(g) subjecting the internally-primed duplexes to conditions for primer extension to produce dicistronic DNA molecules, each containing a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said first polypeptide- or second polypeptide-coding DNA homologs, and said downstream cistron comprising the other of said first polypeptide- or second polypeptide-coding DNA homologs.

5. The method of claim 4 wherein steps (a)-(d) are performed concurrently in one reaction vessel.

6. The method of claim 4 wherein the genes of steps (a) and (c) are present in gene repertoires formed by isolating mRNA from at least 10^3 peripheral blood lymphocytes.

7. The method of claim 6 wherein said repertoire of first polypeptide genes comprises at least 10^3 different first polypeptide genes.

8. The method of claim 6 wherein said repertoire of second polypeptide genes comprises at least 10^3 different second polypeptide genes.

9. The method of claim 4 further comprising step (h) wherein said plurality of different dicistronic DNA molecules is combined with said outside first gene primer and said outside second gene primer to form a third PCR admixture, and subjecting said third PCR admixture to a plurality of PCR thermocycles to produce an amplified library of dicistronic DNA molecules.

10. The method of claim 4 wherein said outside first gene primer hybridizes to a framework, leader or promoter region of a V_H immunoglobulin gene.

11. The method of claim 4 wherein said outside second gene primer hybridizes to a J_L or framework region of a V_L immunoglobulin gene.

12. The method of claim 4 wherein said 3'-terminal priming portion of said inside first gene primer hybridizes to a J_H, hinge or framework region of a V_H immunoglobulin gene.

13. The method of claim 4 wherein said 3'-terminal priming portion of said inside second gene primer hybridizes to a framework, leader or promoter region of a V_L immunoglobulin gene.

14. A method for producing a library of dicistronic DNA molecules, which method comprises:

(a) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) a repertoire of V_H genes,

(ii) a repertoire of V_L genes,

(iii) a V_H PCR primer pair defined by an outside V_H gene primer and an inside V_H gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion; and

(iv) a V_L PCR primer pair defined by an outside V_L gene primer and an inside V_L gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to said 5'-terminal non-priming portion of said inside V_H gene primer, said V_H inside and V_L inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron,

said outside primers being present in said composition in molar excess relative to said inside

primers; and

(b) subjecting said PCR admixture to a plurality of PCR thermocycles, thereby producing said library.

15. The method of claim 14 wherein said signals include a ribosome binding site downstream from said stop codon and a translation initiation codon for said downstream cistron, said translation initiation codon located downstream from said ribosome binding site.

16. The method of claim 15 wherein said translation initiation codon is operatively linked to a polypeptide leader-encoding sequence that is in the same reading frame as the downstream cistron.

17. The method of claim 14 wherein said plurality of PCR thermocycles is at least n+5, wherein n is the number of PCR thermocycles necessary to decrease by a factor of 10 the number of said inside primers by consumption in the formation of inside primer-primed products.

18. The method of claim 17 wherein said repertoire of V_H genes and said outside V_H gene primer are present at a respective molar ratio in the range of $1:10^3$ to $1:10^7$, said repertoire of V_H genes and said inside V_H gene primer are present at a respective molar ratio in the range of $1:10^2$ to $1:10^6$, said repertoire of V_L genes and said outside V_L gene primer are present at a respective molar ratio in the range of $1:10^3$ to $1:10^7$, and said repertoire of V_L genes and said inside V_L gene primer are present at a respective molar ratio in the range of $1:10^2$ to $1:10^6$.

19. The method of claim 17 wherein said repertoire of V_H genes and said outside V_H gene primer are present at a respective molar ratio of about $1:10^4$, said repertoire of V_H genes and said inside V_H gene primer are present at a respective molar ratio of about $1:10^3$, said

repertoire of V_L genes and said outside V_L gene primer are present at a respective molar ratio of about $1:10^4$, and said repertoire of V_L genes and said inside V_L gene primer are present at a respective molar ratio of about $1:10^3$.

20. A method of producing a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, which method comprises:

(a) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) a repertoire of V_H genes,

(ii) a repertoire of V_L genes,

(iii) an outside V_H gene primer

(iv) an outside V_L gene primer, and

(v) a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion, said 3'-terminal priming portion having a nucleotide base sequence complementary to a portion of the primer extension product of one of said outside primers, said 5'-terminal primer template portion having a nucleotide base sequence homologous to a portion of the primer extension product of the other of said outside primers and said cistronic bridge coding portion encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron; and

(b) subjecting said PCR admixture to a plurality of PCR thermocycles, thereby producing said library.

21. A method of producing an isolated dicistronic expression vector capable of expressing V_H and V_L polypeptides from respective V_H- and V_L-coding DNA homologs, said V_H and V_L polypeptide being capable of forming an antibody molecule that binds a preselected antigen, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, a repertoire of V_H genes and a first PCR primer pair defined by an outside V_H gene primer and an inside V_H gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a V_H gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of V_H-coding, double stranded DNA homologs in a double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, a repertoire of V_L genes and a second PCR primer pair defined by an outside V_L gene primer and an inside V_L gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said V_H gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a V_L gene, said V_H inside and V_L inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron and signals necessary for the initiation of translation of the downstream cistron;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of V_L-coding DNA homologs in double stranded form;

(e) separating said double stranded DNA homologs of steps (b) and (d);

(f) hybridizing said separated strands to form a plurality of internally-primed duplexes;

(g) subjecting said internally-primed duplexes to conditions for primer extension to produce a plurality of different dicistronic DNA molecules, each containing a V_H-coding sequence and a V_L-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said V_H- or V_L-coding DNA homologs, and said downstream cistron comprising the other of said V_H-V_L-coding DNA homologs;

(h) operatively linking for expression each of a plurality of said different dicistronic DNA molecules to expression vectors thereby forming a plurality of different V_{HL} expression vectors;

(i) transforming a population of host cells compatible with said expression vector with a plurality of said different V_{HL}-expression vectors to produce a transformed population of host cells whose members contain said V_{HL}-expression vectors;

(j) culturing said transformed population under conditions for expressing the V_H and V_L polypeptides coded for by said dicistronic DNA molecules;

(k) assaying the members of said transformed population for expression of an antibody molecule capable of binding said preselected ligand, thereby identifying transformants containing said dicistronic DNA molecule; and

(l) segregating an identified transformant to step (d) from said population, thereby producing said isolated dicistronic DNA molecule.

22. A kit comprising an enclosure containing, in separate containers, an outside first polypeptide-encoding gene primer, an outside second polypeptide-encoding gene primer, and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer template portion, said 3'-terminal priming portion having a nucleotide base sequence

homologous to a portion of the primer extension product of one of said outside primers, said 5'-terminal primer template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of said outside primers and said cistronic bridge coding portion encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron.

23. The kit of claim 22 wherein said outside first and outside second polypeptide-encoding gene primers are V_H and V_L gene primers, respectively.

24. A kit comprising an enclosure containing, in separate containers, an outside first polypeptide-encoding gene primer, an outside second polypeptide-encoding gene primer, an inside first polypeptide-encoding gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a first polypeptide-encoding gene, and an inside second polypeptide-encoding gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said inside first polypeptide-encoding gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a second polypeptide-encoding gene, said first polypeptide inside and second polypeptide inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron.

25. The kit of claim 24 wherein said outside first and outside second polypeptide-encoding gene primers are V_H and V_L gene primers, respectively.

26. A library of dicistronic DNA molecules produced by the method of any one of Claims 1-3, or 14-20.

27. A dicistronic DNA molecule produced by the method of any one of Claims 4-13.

28. An isolated dicistronic expression vector produced by the method of Claim 21.

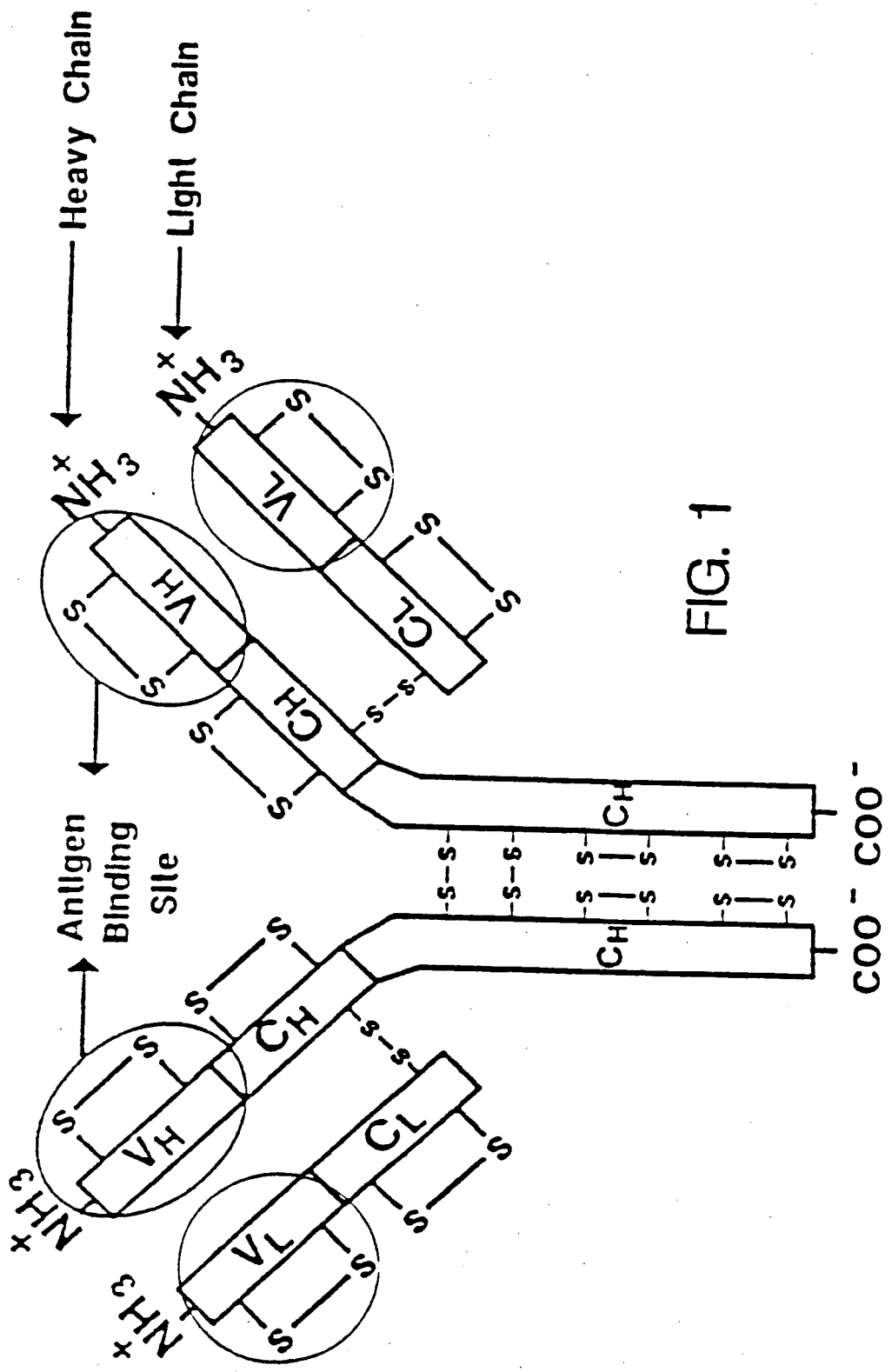


FIG. 1

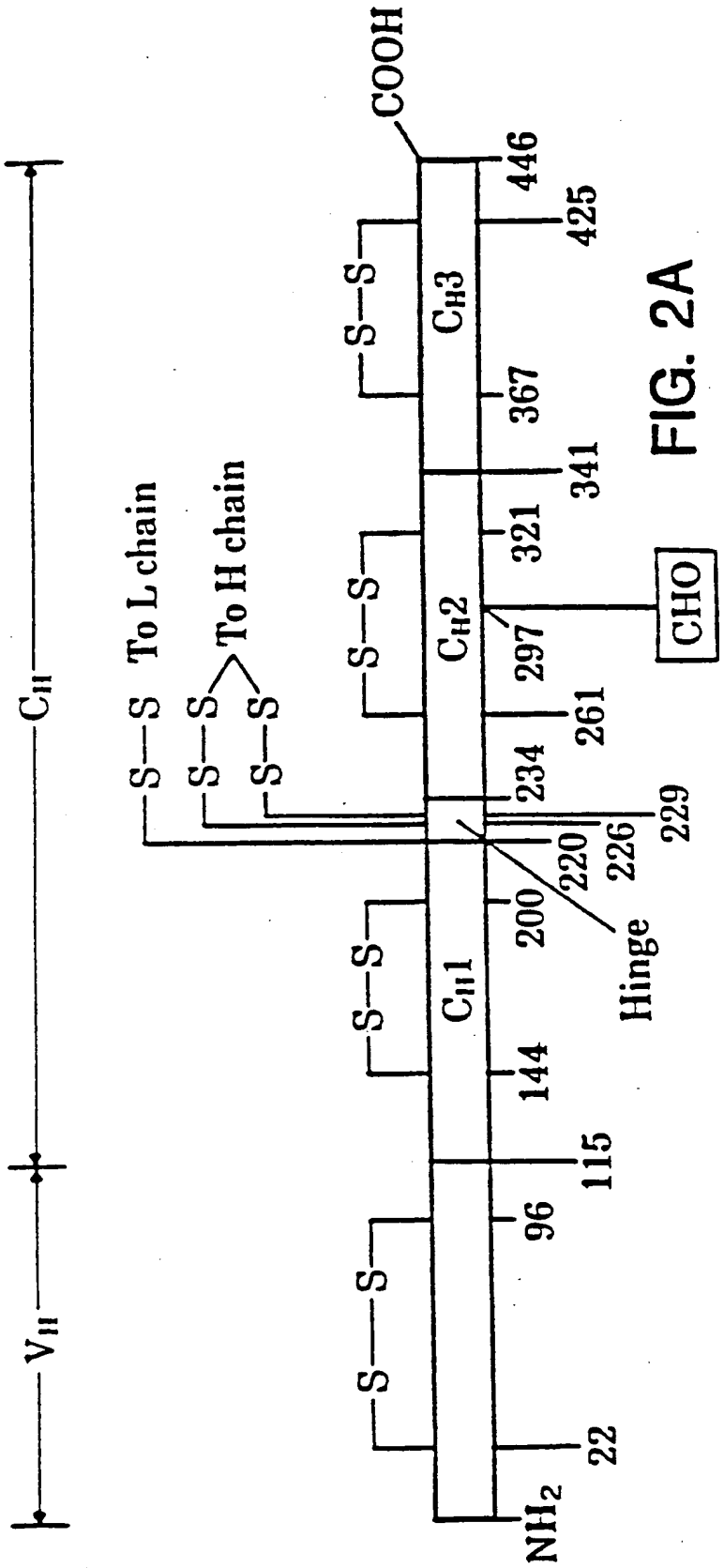


FIG. 2A

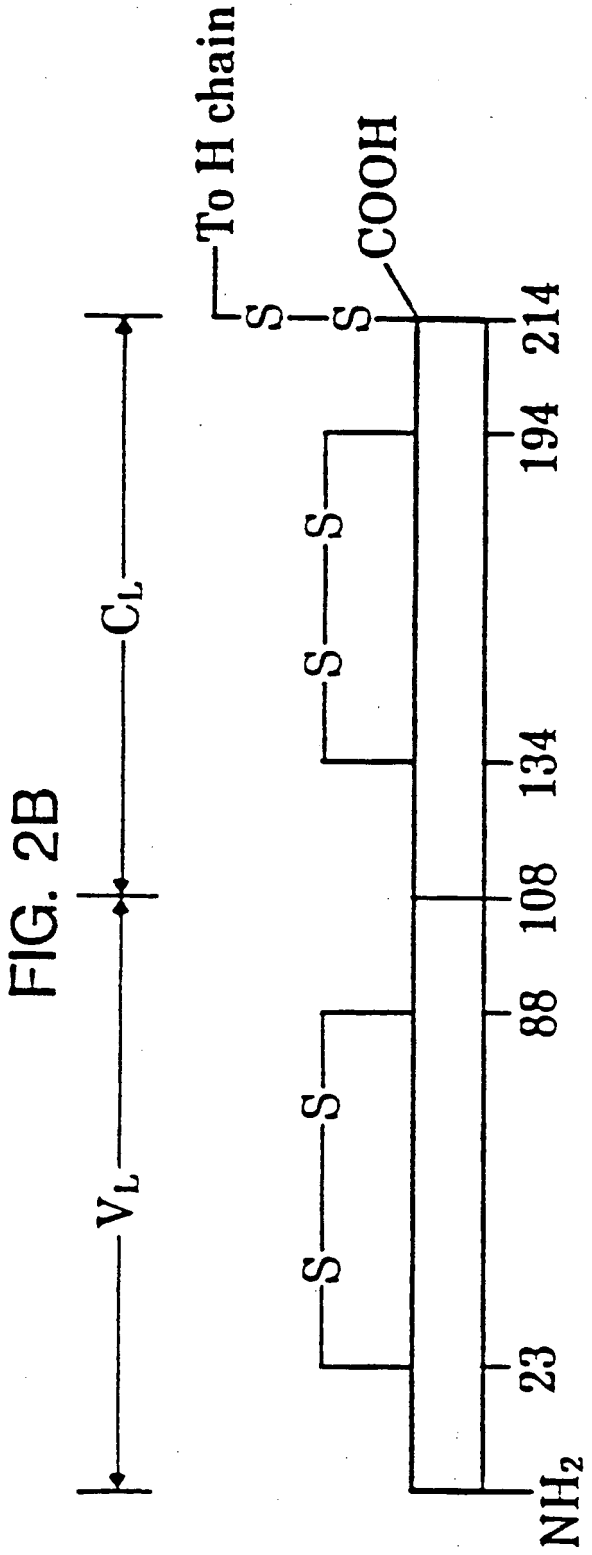


FIG. 2B

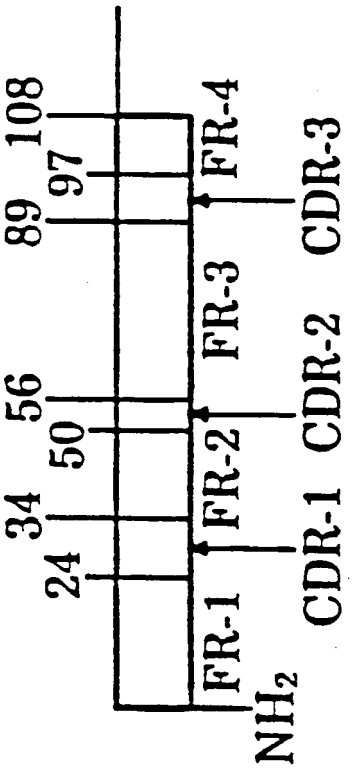


FIG. 2C

```
250  TGGAAAGAGGGCGATAAGAGCTTCTATTTCGACACTAACGTGGCCTATTCCGTCGCACAACAG
     TrpLysGluGlyAspLysSerPheTyrPheAspThrAsnValAlaTyrSerValAlaGlnGln  71

313  AATGACTGGGAAGCTACCGATCCGGCCTTCCGTGAAGCAAACGTGCAGGGTAAAAACCTGATC
     AsnAspTrpGluAlaThrAspProAlaPheArgGluAlaAsnValGlnGlyLysAsnLeuIle  92

376  GAATGGCTGCCAGGCTCCACCATCTGGGCAGGTAAGCGCTTCTACCAACGTCATGACGTTCAT
     GluTrpLeuProGlySerThrIleTrpAlaGlyLysArgPheTyrGlnArgHisAspValHis  113

439  ATGATCGACTTCTACTACTGGGATATTTCTGGTCCTGGTGCCGGTCTGGAAAACATCGATGTT
     MetIleAspPheTyrTyrTrpAspIleSerGlyProGlyAlaGlyLeuGluAsnIleAspVal  134

502  GGCTTCGGTAAACTCTCTCTGGCAGCAACCCGCTCCTCTGAAGCTGGTGGTTCTTCCTCTTTC
     GlyPheGlyLysLeuSerLeuAlaAlaThrArgSerSerGluAlaGlyGlySerSerSerPhe  105

565  GCCAGCAACAATATTTATGACTATACCAACGAAACCGCGAACGACGTTTTCGATGTGCGTTTA
     AlaSerAsnAsnIleTyrAspTyrThrAsnGluThrAlaAsnAspValPheAspValArgleu  176

628  GCGCAGATGGAAATCAACCCGGGC
     AlaGlnMetGluIleAsnProGly                                         184
```

FIGURE 3

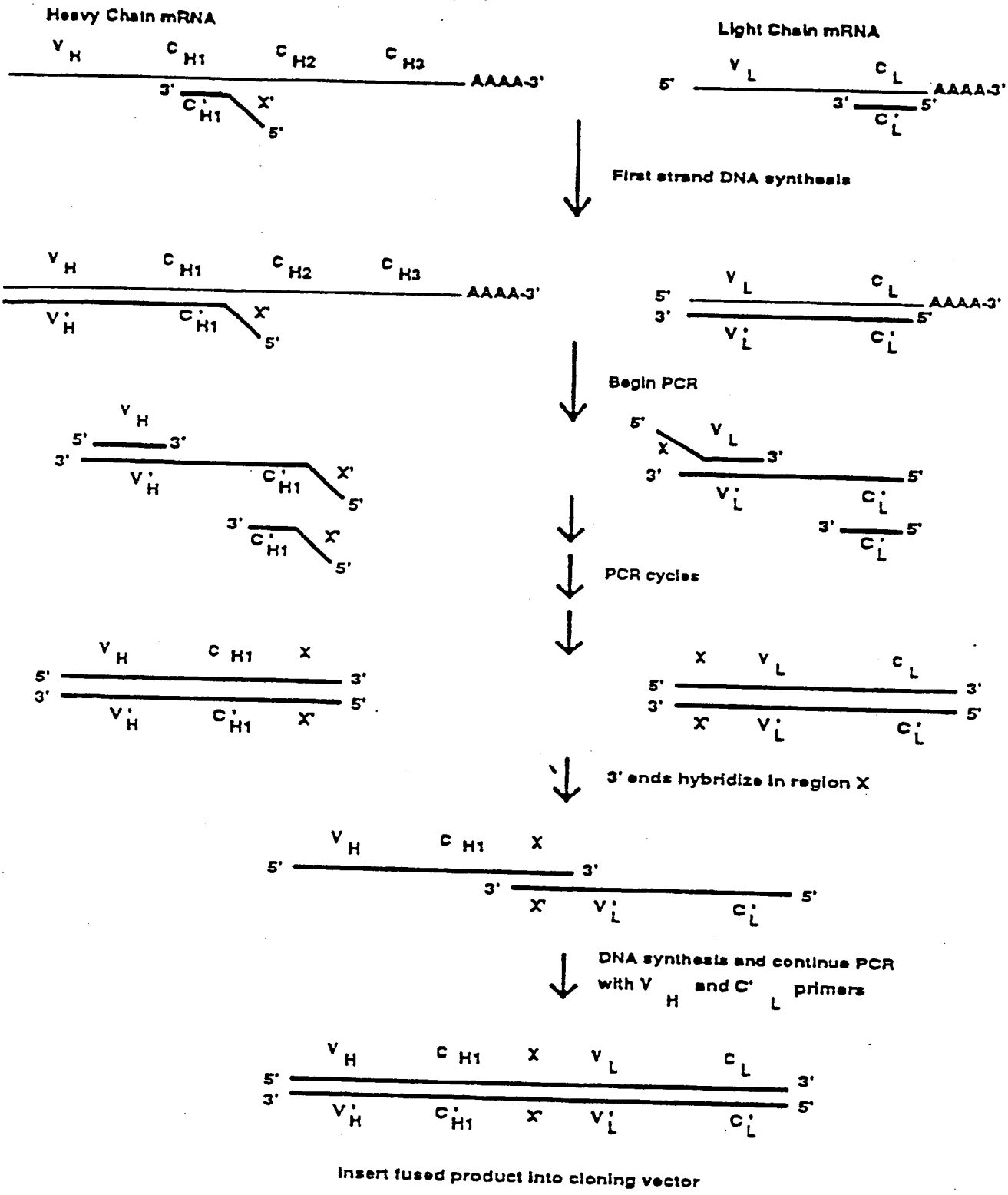


FIGURE 4

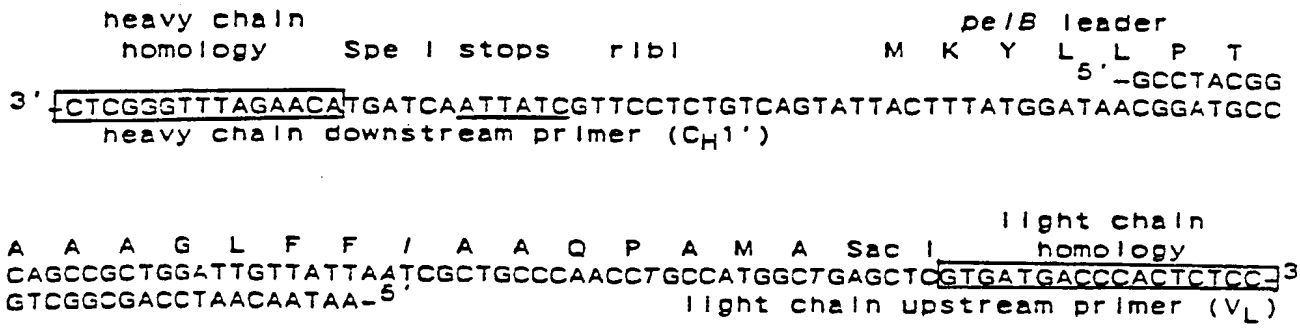


FIGURE 5

V_H EXPRESSION VECTOR:

Not I Ribosome Binding Site

```
5' GGCCGCAAATTCTATTTCAAGGAGACAGTCATA
     CGTTTAAGATAAAGTTCCTCTGTCAGTAT
```

Pel B Leader

```
MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA
```

NcoI XhoI XbaI SpeI

V_H backbone

```
AlaGlnProAlaMetAlaGlnValLysLeuLeuGlu Thr
GCCCAACCAGCCATGGCCCAGGTGAAACTGCTCGAGATTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACTTTGACGAGCTCTAAAGATCTGA
```

EcoRI

```
SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT
```

FIGURE 6A

V. EXPRESSION VECTOR:

```
          EcoRI                    Ribosome Binding Site

5'      TGAATTCTAAACTAGTCGCCAAGGAGACAGTCATA
3' TCGAACTTAAGATTTGATCAGCGGTTCCTCTGTCAGTAT

                      Pel B Leader

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeu
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTC
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAG

               NcoI      SacI

AlaAlaGlnProAlaMetAlaGluLeu
GCTGCCCAACCAGCCATGGCCGAGCTC
CGACGGGTTGGTCGGTACCGGCTCGAG


       XbaI
         Stop  Stop
GTCAGTTCTAGAGTTAAGCGGCCG
CAGTCAAGATCTCAATTCGCCGGCAGCT
```

FIGURE 6B

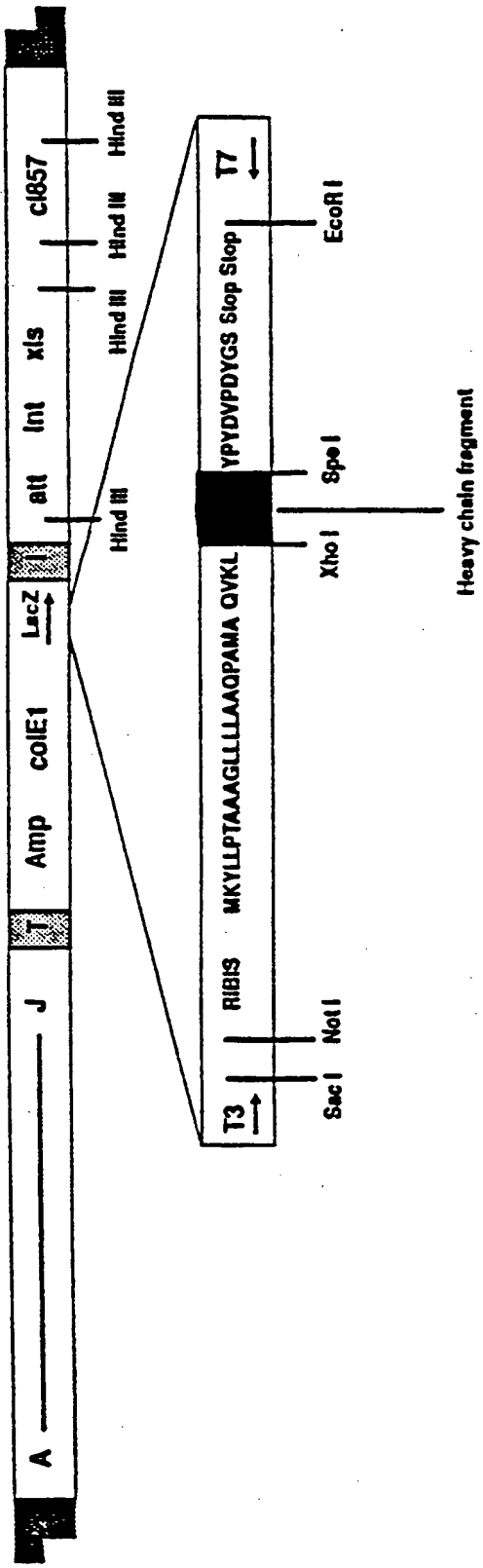


FIGURE 7

MODIFIED V_L EXPRESSION VECTOR:

```
          Not I        Ribosome Binding Site

5' GAGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCATA
3'      CGCCGGCGTTTAAGATAAAGTTCCTCTGTCAGTAT

                  Pel B Leader

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA

             NcoI                XhoI    XbaI SpeI

AlaGlnProAlaMetAlaGlnValGlnLeuLeuGlu         Thr
GCCCAACCAGCCATGGCCCAGGTGCAGCTGCTCGAGATTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACGTCGACGAGCTCTAAAGATCTGA

                                          EcoRI
SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT
```

FIGURE 8A

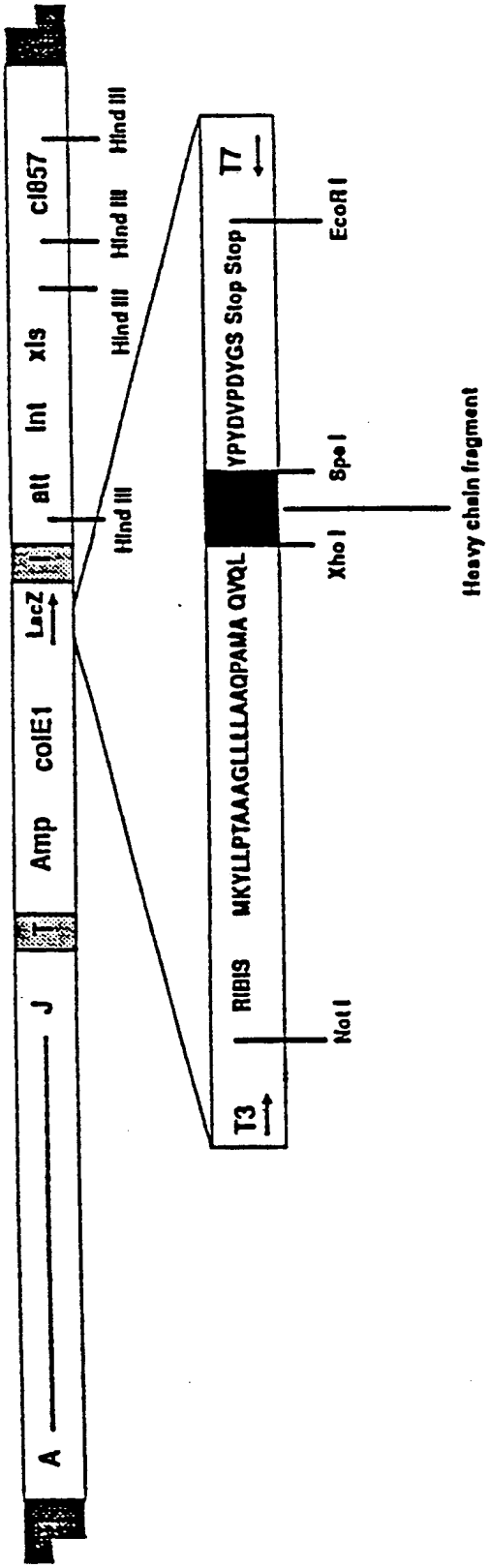


FIGURE 8B

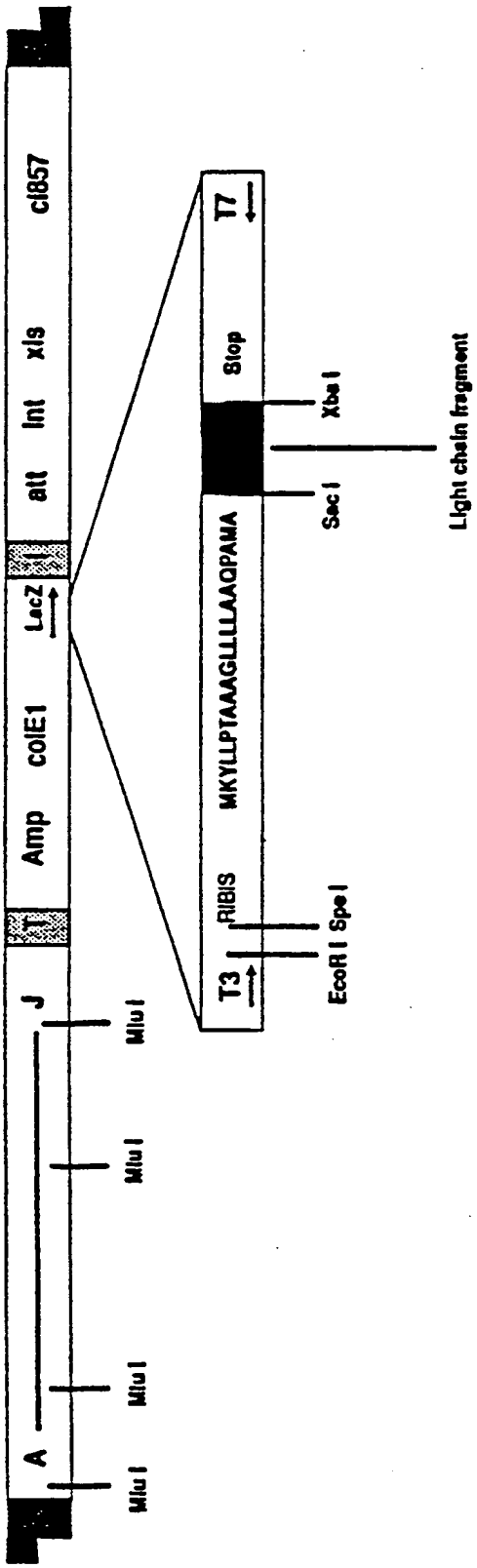


FIGURE 9

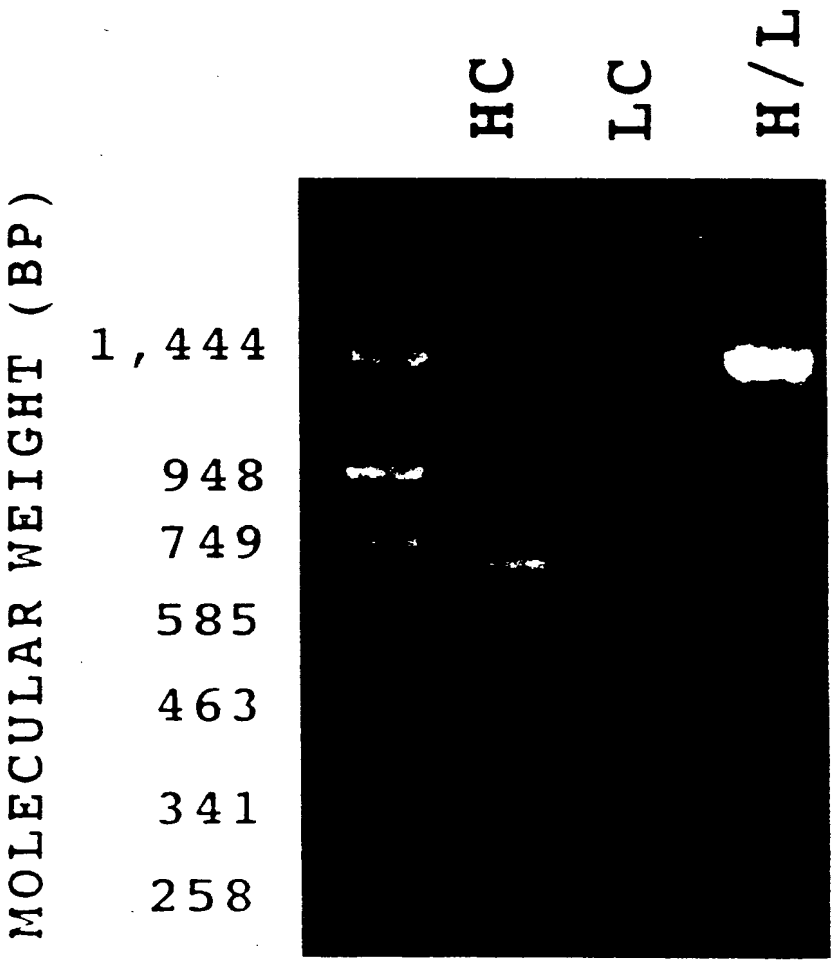


FIG. 10

FIG. 11

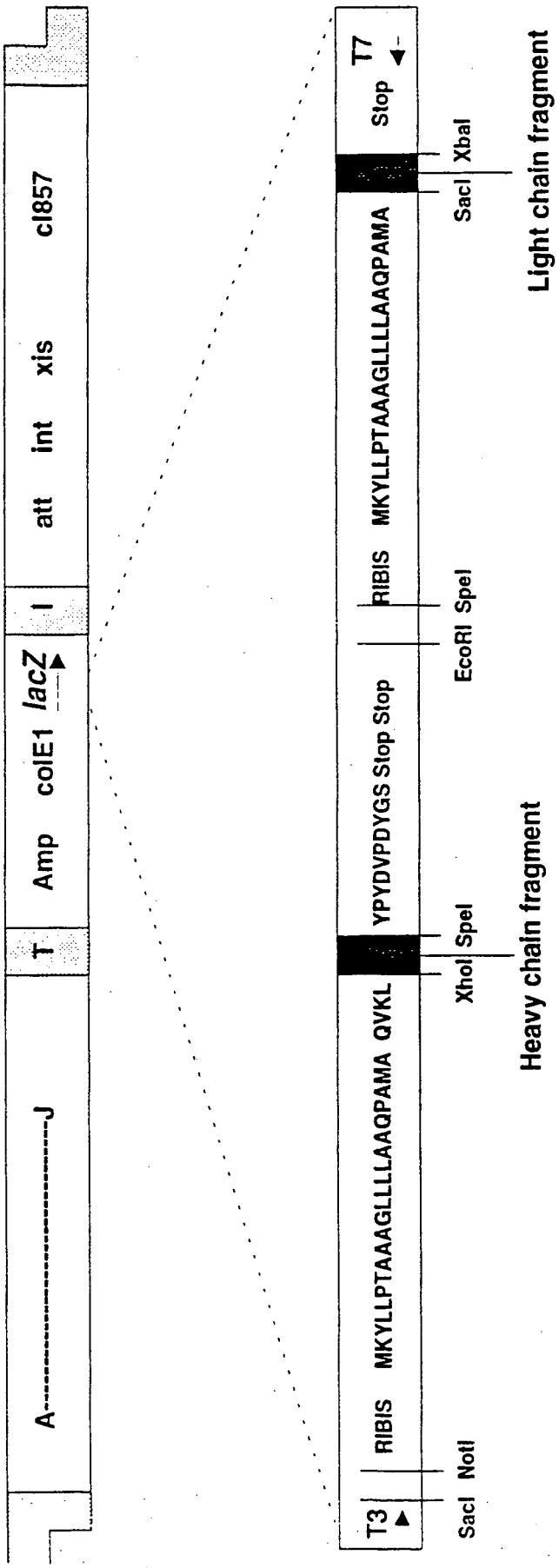


FIGURE 12

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/01475

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/10; C12N15/13; C12Q1/68; //
C12N15/62

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C12Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
P,X	FASEB JOURNAL. vol. 5, no. 6, 19 March 1991, BETHESDA, MD US; A-1717, ABSTR. 7820 R. L. MULLINAX ET AL.: 'ANTIBODY EXPRESSION LIBRARIES IN E. COLI: SIMPLIFIED CONSTRUCTION USING PCR-MEDIATED GENE FUSION' see abstract & 75TH ANNUAL MEETING OF THE FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY, ATLANTA, GEORGIA, USA, APRIL 21-25, 1991 ---	1-28
A	WO,A,9 014 430 (SCRIPPS CLINIC AND RESEARCH FOUNDATION) 29 November 1990 ---	

-/--

* Special categories of cited documents :[10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
09 JUNE 1992	12 JUN 1992
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	THIELE U.H.-C.H.

Form PCT/ISA/210 (second sheet) (January 1985)

International Application No.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	GENE. vol. 77, 1989, AMSTERDAM NL pages 61 - 68; R. M. HORTON ET AL.: 'Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension'	

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9201475 SA 58103

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 09/06/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9014430	29-11-90	AU-A- 5673390	18-12-90
		AU-A- 5813890	18-12-90
		CA-A- 2016841	16-11-90
		CA-A- 2016842	16-11-90
		EP-A- 0472638	04-03-92
		EP-A- 0425661	08-05-91
		JP-T- 4500607	06-02-92
		WO-A- 9014424	29-11-90
		AU-A- 5834490	18-12-90
		EP-A- 0478627	08-04-92
		WO-A- 9014443	29-11-90

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/10, 15/13, C12Q 1/68
// C12N 15/62

A1

(11) International Publication Number: WO 92/15678

(43) International Publication Date: 17 September 1992 (17.09.92)

(21) International Application Number: PCT/US92/01475

(22) International Filing Date: 27 February 1992 (27.02.92)

(30) Priority data:
663,442 1 March 1991 (01.03.91) US

(71) Applicant: STRATAGENE [US/US]; 11099 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventor: SORGE, Joseph, A. ; P.O. Box 9437, Rancho Santa Fe, CA 92067 (US).

(74) Agents: SIMPSON, Andrew, H. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, Suite 1600, Newport Beach, CA 92660 (US).

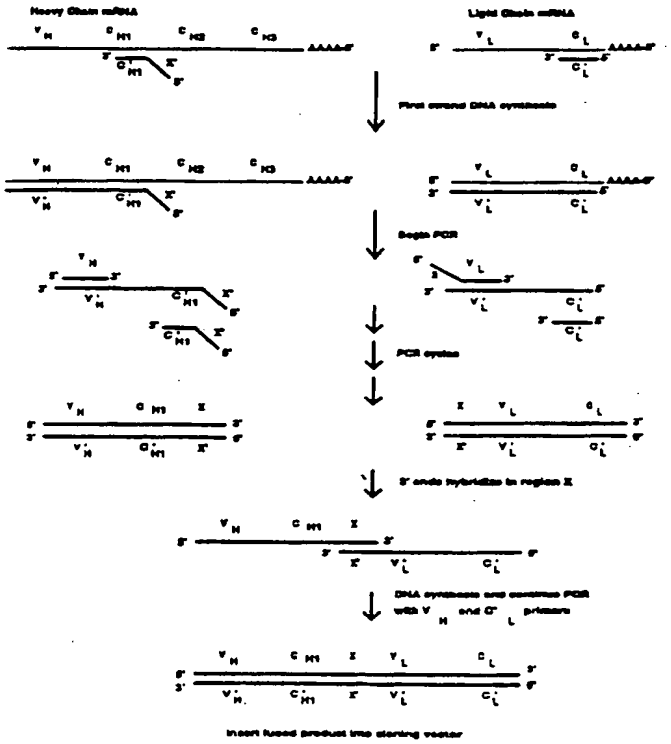
(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: PCR GENERATED DICISTRONIC DNA MOLECULES FOR PRODUCING ANTIBODIES

(57) Abstract

A method of producing dicistronic DNA molecules each having upstream and downstream cistrons respectively coding for the first and second polypeptides of a heterodimeric receptor. Kits including, in separate containers, the primers and/or vectors of the invention in amounts sufficient to produce and/or express the dicistronic DNA molecules.



* (Referred to in PCT Gazette No. 27/1992, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

PCR GENERATED DICISTRONIC DNA MOLECULES FOR PRODUCING ANTIBODIES

Technical Field

The present invention relates to a method for producing a library of dicistronic DNA molecules useful in expressing heterodimeric receptors, such as antibodies, T cell receptors and the like.

Background

The expression of antibody libraries in bacteria has opened up new ways to uncover monoclonal antibody specificities. The antigen binding domain of the antibody is composed of a heavy and a light chain. These chains are each encoded by separate genes. To reconstruct a complete binding domain in bacteria, both heavy and light chain coding sequences are typically coexpressed, which involves two cloning steps, one for the heavy chain and one for the light. This is generally accomplished by either inserting both heavy and light chain coding sequences into one vector, or by first making separate heavy and light chain libraries and recombining the genomes to make a combinatorial library encoding random combinations of the heavy and light sequences. In either case, the need to clone two separate DNA fragments is cumbersome and, therefore, a method that could fuse both heavy and light chain sequences together prior to vector ligation would be desirable.

Brief Description of the Invention

The present invention contemplates a method of producing dicistronic DNA molecules each having upstream and downstream cistrons respectively coding for first and second polypeptides of a heterodimeric protein, such as a receptor. The method comprises the

following steps:

(A) Forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, first polypeptide-encoding genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer. The inside first gene primer has a 3'-terminal priming portion and, preferably, a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a first gene.

(B) Subjecting the first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form.

(C) Forming a second PCR admixture by combining, in a PCR buffer, second polypeptide-encoding genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer. The inside gene primer has a 3'-terminal priming portion and, preferably, a 5'-terminal hybridizing portion complementary to a hybridizable portion of the 5'-terminal non-priming portion of the first inside gene primer. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a second polypeptide-coding gene.

The first and second inside primers, when hybridized, form a duplex that codes for a double-stranded cistronic bridge that links the upstream and downstream cistrons. One strand of the bridge codes for (i) at least one stop codon in the same reading frame as said upstream cistron, (ii) signals for the initiation of translation of the downstream in cistron. Preferably, such signals include a ribosome binding site downstream from the stop codon, and at

least one translation initiation codon in the same reading frame as the downstream cistron, the initiation codon being located downstream from the ribosome binding site.

(D) Subjecting the second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form.

(E) Separating the double stranded DNA homologs produced in steps (B) and (D).

(F) Hybridizing the separated strands of step (E) to form internally-primed duplexes.

(G) Subjecting the internally-primed duplexes to conditions for primer extension to produce a dicistronic DNA molecule. Each of the dicistronic DNA molecules produced contains a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by the cistronic bridge. The upstream cistron comprises one of the first polypeptide- or second polypeptide-coding DNA homologs. The downstream cistron comprises the other of the first polypeptide- or second polypeptide-coding DNA homologs.

Preferably, steps (A)-(D) are performed concurrently in one reaction vessel.

Preferably, the polypeptide-encoding genes of steps (A) and (B) are present in respective repertoires of conserved genes. When used, the repertoires of steps (A) and (C) are usually formed by isolating mRNA from at least about 10^3, preferably at least about 10^7 lymphocytes. It is preferred that the repertoire of first polypeptide genes comprises at least 10^5 different first polypeptide genes, and that the repertoire of second polypeptide genes comprises at least 10^5 different second polypeptide genes.

However, it should be noted that the method of the present invention can be used to operatively link for polycistronic expression any two genes. Thus, this invention can be used to physically link two genes from a single cell, such as a B cell, T cell, and the like, and thereby take advantage of a native immune system's ability to select operative gene pairs from the immunological repertoire. Similarly, operative gene pairs, i.e., a pair of genes encoding a heterodimeric receptor, from cells such as hybridomas, quadromas and the like, can be physically linked using the method of this invention.

Preferably the method further comprises step (H) wherein the dicistronic DNA molecules are PCR amplified by combining them with the outside first gene primer and the outside second gene primer to form a third PCR admixture. The third PCR admixture is then subjected to a plurality of PCR thermocycles. When a repertoire of first and/or second polypeptide-encoding genes is used, an amplified library of dicistronic DNA molecules is produced.

In preferred embodiments, the amplified products of step (H) are operatively linked for expression to a vector, preferably a phage vector. Preferably, the steps for operatively linking the dicistronic DNA molecules to a vector and isolating a recombinant vector that expresses a desired heterodimeric receptor include the following:

(i) Preparation of vector DNA and the dicistronic DNA molecules by cleavage with appropriate restriction enzyme(s) to form cohesive termini.

(ii) Ligation of the digested vector with the dicistronic DNA molecules via the cohesive termini.

(iii) Packaging of the ligated DNA (rDNA) into bacteriophage particles that can form plaques on appropriate bacterial hosts.

(iv) Identification of recombinant bacteriophages carrying the desired dicistronic DNA molecules.

(v) Plaque purification of selected recombinant bacteriophages.

Where the heterodimeric receptor is an antibody, the outside first gene primer hybridizes to a framework, leader or promoter region of a V_H immunoglobulin gene, and the outside second gene primer hybridizes to a J_L, constant or framework region, of a V_L immunoglobulin gene. The 3'-terminal priming portion of the inside first gene primer hybridizes to a J_H, hinge, constant, or framework region of a V_H immunoglobulin gene, and the 3'-terminal priming portion of the inside second gene primer hybridizes to a framework, leader or promoter region of a V_L immunoglobulin gene.

In another embodiment, a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, is produced by the following steps:

(A) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) V_H genes,

(ii) V_L genes,

(iii) an outside V_H gene primer

(iv) an outside V_L gene primer, and

(v) a linking primer having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer

extension product of one of the outside primers. The 5'-terminal primer template portion has a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

(B) Subjecting the PCR admixture of step (A) to a plurality of PCR thermocycles.

In preferred embodiments, the method further comprises steps (C)-(H) as follows:

(C) Subjecting the internally-primed duplexes to conditions for primer extension to produce dicistronic DNA molecules, each containing a V_H-coding sequence and a V_L-coding sequence linked by the cistronic bridge. The upstream cistron comprises one of the V_H- or V_L-coding DNA homologs, and the downstream cistron comprising the other of the V_H- V_L-coding DNA homologs.

(D) Operatively linking for expression the different dicistronic DNA molecules produced in step (C) to expression vectors, preferably phage vectors, thereby forming a plurality of V_{HL} expression vectors.

(E) Transforming a population of host cells, preferably E. coli compatible with the expression vector with a plurality of the V_{HL}-expression vectors to produce a transformed population of host cells whose members contain the V_{HL}-expression vectors.

(F) Culturing the transformed population under conditions for expressing the V_H and V_L polypeptides coded for by the dicistronic DNA molecules.

(G) Assaying the members of the transformed population for expression of an antibody molecule capable of binding a preselected ligand, thereby identifying transformants containing the dicistronic

SUBSTITUTE SHEET

DNA molecule.

(H) Segregating an identified transformant in step (G) from the population, thereby producing the isolated dicistronic DNA molecule.

Also contemplated are kits for producing a dicistronic DNA molecule as described herein. In one embodiment, the kit is an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, and a linking primer defining a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

Another contemplated kit comprises an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, an inside first polypeptide, preferably a V_H, gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a V_H gene. The kit also contains an inside second polypeptide, preferably a V_L, gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of the first polypeptide gene primer, the 3'-terminal priming portion of which comprises a

nucleotide sequence homologous to a conserved portion of a second polypeptide gene. The first polypeptide inside and second polypeptide inside primers, when hybridized, form a duplex that codes for a double-stranded DNA molecule containing the before described cistronic bridge for linking the upstream and downstream cistrons.

Brief Description of the Drawings

Figure 1 illustrates the principal structural features of an immunoglobulin molecule. The circled areas on the heavy and light chains represent the variable regions, (V_H) and (V_L), a heterodimeric polypeptide containing a biologically active (ligand binding) portion of that region, and genes coding for the individual polypeptides, are produced by the methods of the present invention.

Figure 2 contains three panels. Panel 2A illustrates various features of the heavy chain of human IgG (IgG1 subclass). Numbering is from the N-terminus on the left to the C-terminus on the right. Note the presence of four domains, each containing an intrachain disulfide bond (S-S) and spanning approximately 110 amino acid residues. The symbol CHO stands for carbohydrate. The V region of the heavy (H) chain (V_H) resembles V_L in having three hypervariable complementarity determining regions (CDR'S) (not shown).

Panel 2B and 2C illustrate various features of a human kappa (K) chain. Numbering is from the N-terminus on the left to the C-terminus on the right. Note in Panel 2B the intrachain disulfide bond (S-S) spanning about the same number of amino acid residues in the V_L and C_L domains. Panel 2C shows the locations of the CDRs in the V_L domain. Segments

SUBSTITUTE SHEET

outside the CDR are the framework segments (FR).

Figure 3 illustrates a portion of the nucleotide base sequence of the 1661 base pair gene lamB sequence from residue number 250 to residue number 651. The base sequences are shown conventionally from left to right and in the direction of 5' terminus to 3' terminus using the single letter nucleotide base code (A = adenine, T = thymine, C = cytosine and G = guanine). The position of the nucleotide base sequence is indicated by the numbers in the left margin of the figure.

The reading frame of the structural lamB gene is indicated by placement of the deduced amino acid residue sequence of the lambda receptor protein for which it codes below the nucleotide sequence such that the triple letter code for each amino acid residue is located directly below the three bases (codon) coding for each residue. The residue sequence is shown conventionally from left to right and in the direction of amino terminus to carboxy terminus. The position of the amino acid residue sequence is indicated by the numbers in the right margin of the figure.

Figures 4A and 4B illustrates the strategy used to create immunoglobulin heavy and light chain PCR fusion products. RNA and DNA are represented by dotted and solid lines, respectively. Regions of the immunoglobulin heavy chain coding strand area designated V_H, C_H1, C_H2, and C_H3 correspond to those functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime (') following the symbol. Regions V_L and C_L are similarly labelled for the light chain. A region, X, unrelated to the natural immunoglobulin sequences is introduced into the fusion product by attaching X to the 5' ends of the C_H1' inside and V_L inside primers.

Throughout the ensuing detailed description when reference is made to "Figure 4," it is meant to refer to Figures 4A and 4B.

Figure 5 illustrates human fusion PCR inside primers. The heavy chain C_H1' inside primer sequence is written 3' to 5' and the light chain V_L inside primer sequence is written 5' to 3'. Note that it is not the primer strands that cross-prime to create the fusion molecule, but the complementary PCR product strands. Boxed nucleotides represent regions where the C_H1' primer hybridizes to the 3' end of C_H1 on human IgG heavy chain mRNA or where the V_L primer hybridizes to the 5' end of V_L framework-1 on human kappa light chain cDNA. Underlined sequences indicate the two stop codons. The italicized amino acid and nucleotides indicate changes in sequence from the original pelB leader sequence. The mouse fusion-PCR internal primers overlap in a similar manner.

Figure 6 illustrates the sequences of the synthetic DNAs inserted into Lambda ZAP to produce Lambda Zap II V_H (ImmunoZAP H) (Panel A) and Lambda Zap V_L (ImmunoZAP L) (Panel B) expression vectors. The various features required for these vectors to express the V_H and V_L-coding DNA homologs include the Shine-Dalgarno ribosome binding site, a leader sequence to direct the expressed protein to the periplasm as described by Mouva et al., J. Biol. Chem., 255:27, 1980, and various restriction enzyme sites used to operatively link the V_H and V_L homologs to the expression vector. The V_H expression-vector sequence also contains a short nucleic acid sequence that codes for amino acids typically found in variable regions heavy chain (V_H Backbone). This V_H Backbone is just upstream and in the proper reading as the V_H DNA homologs that are operatively linked into the Xho I and Spe I restriction sites. The V_L DNA homologs are operatively linked into the V_L sequence (Panel B) at the Sac I and Xba I restriction enzyme sites.

Figure 7 illustrates the major features of the bacterial expression vector Lambda Zap II V_H (ImmunoZAP H) (V_H- expression vector). The amino acids encoded by the synthetic DNA sequence from Figure 6A is shown at the top along with the T_3 polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The V_H DNA homologs were inserted into the phagemid that is produced by the in vivo excision protocol described by Short et al., Nucleic Acids Res., 16:7583-7600, 1988. The V_H DNA homologs were inserted into the Xho I and Spe I restriction enzyme sites. The read through transcription produces the decapeptide epitope (tag) that is located just 3' of the cloning sites.

Figure 8 illustrates, in Panels 8A and 8B, the major features of the bacterial expression vector Lambda ZAP II Modified V_H (Modified ImmunoZAP H) (V_H- expression vector) (IZ H). The amino acids encoded by the synthetic DNA sequence from Panel 8A is shown along with the T_3 polymerase promoter from Lambda ZAP II. The orientation of the insert in Lambda ZAP II is as presented. The insert was modified by the elimination of the Sac I site between the T_3 polymerase and Not I site and by the change of amino acids at the 5' end of the heavy chain from QVKL to QVQL (a lysine residue was changed to a glutamine residue). The V_H and V_L DNA homologs were inserted into the Xho I and Xba I cloning sites of the phagemid as described in Figure 7 and shown in Panel 8B. The modifications were made to create a fusion-PCR library from hybridoma RNA, to overcome decreased efficiency of secretion of positively charged amino acids in the amino terminus of the protein. Inouye et al., Proc. Natl. Acad. Sci., USA, 85;7685-7689 (1988), and to

make the V_L Sac I cloning site a unique restriction site.

Figure 9 illustrates the major features of the bacterial expression vector Lambda Zap II V_L (ImmunoZAP L) (V_L expression vector). The amino acids encoded by the synthetic DNA sequence shown in Figure 6B is shown at the top along with the T_3 polymerase promoter from Lambda Zap II. The orientation of the insert in Lambda Zap II is as presented. The V_L DNA homologs are inserted into the Sac I and Xba I cloning sites of the phagemid as described in Figure 7.

Figure 10 illustrates an ethidium bromide stained agarose gel. After PCR amplification from human cloned DNA of heavy chain alone (HC), light chain alone (LC), and the heavy/light dicistronic DNA molecule (H/L), DNA samples were electrophoresed. The expected sizes of the HC, LC, and H/L products visualized on the gel were approximately 730, 690, and 1,390 base pairs, respectively.

Figure 11 illustrates an autoradiogram showing signals obtained from human phage clones. Approximately 100 lambda phage were spotted onto E. coli lawns, creating plaques that were overlaid with nitrocellulose filters previously soaked in 10 mM isopropylbeta-D-thiogalactopyranoside (IPTG) to induce Fab expression. Following overnight incubation, the filters were reacted with ^{125}I-tetanus toxoid probe. After washing, the filters were exposed to X-ray film. The column on the right represents the parental clones that were selected from a combinatorial library. Mullinax et al., Proc. Natl. Acad. Sci., USA, 87:8095-8099 (1990). The column on the left represents clones that were generated by amplifying, the combinatorial lambda clone DNA with the V_H and C_L' outside primers, C_H1' and V_L inside primers, followed by recloning in

the modified ImmunoZAP H vector. Clone 7G1 is a negative control which expresses an Fab that does not react with tetanus toxoid. Clones 10C1 and 6C1 both produce Fabs that react with tetanus toxoid. IZ H is the modified heavy chain ImmunoZAP H vector without an insert.

Figure 12 illustrates the major features of the bacterial expression vector lambda ZAP H/L (ImmunoZAP H/L) (combined V_H- and V_L-expression vector). The ImmunoZAP H/L vector is created from the heavy and light chain libraries by fusing the vectors at the Eco R1 site. DNA is purified from the light chain library and restriction digested with Mlu 1 and Eco R1. This cleaves the DNA from the left arm of the vector into several pieces while leaving the right arm with the light chain inserts intact. DNA is purified from the heavy chain libraries and restriction digested with Hind III and Eco R1. This cleaves the DNA from the right arm of the vector into several pieces while leaving the left arm with the heavy chain inserts intact. The intact left arm of the heavy chain vector containing the heavy chain inserts and right arm of the light chain vector containing the light chain inserts are then mixed and ligated at the common Eco R1 restriction site.

Detailed Description of the Invention

A. Definitions

Nucleotide: A monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3'

or 5' position of the pentose it is referred to as a nucleotide.

Base Pair (bp): A partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double stranded DNA molecule. In RNA, uracil (U) is substituted for thymine.

Nucleic Acid: A polymer of nucleotides, either single or double stranded.

Gene: A nucleic acid whose nucleotide sequence codes for an RNA or polypeptide. A gene can be either RNA or DNA.

Complementary Bases: Nucleotides that normally pair up when DNA or RNA adopts a double stranded configuration.

Complementary Nucleotide Sequence: A sequence of nucleotides in a single-stranded molecule of DNA or RNA that is sufficiently complementary to that on another single strand to specifically hybridize to it with consequent hydrogen bonding.

Conserved: A nucleotide sequence is conserved with respect to a preselected (reference) sequence if it non-randomly hybridizes to an exact complement of the preselected sequence.

Hybridization: The pairing of substantially complementary nucleotide sequences (strands of nucleic acid) to form a duplex or heteroduplex by the establishment of hydrogen bonds between complementary base pairs. It is a specific, i.e. non-random, interaction between two complementary polynucleotides that can be competitively inhibited.

Nucleotide Analog: A purine or pyrimidine nucleotide that differs structurally from A, T, G, C, or U, but is sufficiently similar to substitute for the normal nucleotide in a nucleic acid molecule.

DNA Homolog: Is a nucleic acid having a preselected conserved nucleotide sequence and a sequence coding for a receptor capable of binding a preselected ligand.

Receptor: A receptor is a molecule, such as a protein, glycoprotein and the like, that can specifically (non-randomly) bind to another molecule.

Antibody: The term antibody in its various grammatical forms is used herein to refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v).

Antibody Combining Site: An antibody combining site is that structural portion of an antibody molecule comprised of a heavy and light chain variable and hypervariable regions that specifically binds (immunoreacts with) an antigen. The term immunoreact in its various forms means specific binding between an antigenic determinant-containing molecule and a molecule containing an antibody combining site such as a whole antibody molecule or a portion thereof.

Monoclonal Antibody: The phrase monoclonal antibody in its various grammatical forms refers to a population of antibody molecules that contains only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore

contain an antibody molecule having a plurality of antibody combining sites, each immunospecific for a different antigen, e.g., a bispecific monoclonal antibody.

Upstream: In the direction opposite to the direction of DNA transcription, and therefore going from 5' to 3' on the non-coding strand, or 3' to 5' on the mRNA.

Downstream: Further along a DNA sequence in the direction of sequence transcription or read out, that is traveling in a 3'- to 5'-direction along the non-coding strand of the DNA or 5'- to 3'-direction along the RNA transcript.

Cistron: Sequence of nucleotides in a DNA molecule coding for an amino acid residue sequence.

Stop Codon: Any of three codons that do not code for an amino acid, but instead cause termination of protein synthesis. They are UAG, UAA and UGA. Also referred to as a nonsense or termination codon.

Leader Polypeptide: A short length of amino acid sequence at the amino end of a protein, which carries or directs the protein through the inner membrane and so ensures its eventual secretion into the periplasmic space and perhaps beyond. The leader sequence peptide is commonly removed before the protein becomes active.

Reading Frame: Particular sequence of contiguous nucleotide triplets (codons) employed in translation. The reading frame depends on the location of the translation initiation codon.

Inside Primer: An inside primer is a polynucleotide that has a priming region located at the 3' terminus of the primer which typically consists

of 15 to 30 nucleotide bases. The 3' terminal-priming portion is capable of acting as a primer to catalyze nucleic acid synthesis. The 5'-terminal priming portion comprises a non-priming portion.

Outside Primer: An outside primer comprises a 3'-terminal priming portion and a portion that may define an endonuclease restriction site which is typically located in a 5'-terminal non-priming portion of the outside primer.

B. Methods

The present invention contemplates a method of isolating from a repertoire of conserved genes a pair of genes coding for a dimeric receptor having a preselected activity. Preferably, the receptor will be a heterodimeric polypeptide capable of binding a ligand, such as an antibody molecule or immunologically active portion thereof, a cellular receptor, or a cellular adhesion protein coded for by one of the members of a family of conserved genes, i.e., genes containing a conserved nucleotide sequence of at least about 10 nucleotides in length.

Exemplary conserved gene families encoding different polypeptide claims of a dimeric receptor are those coding for immunoglobulins, major histocompatibility complex antigens of class I or II, lymphocyte receptors, integrins and the like.

A gene can be identified as belonging to a repertoire of conserved genes using several methods. For example, an isolated gene may be used as a hybridization probe under low stringency conditions to detect other members of the repertoire of conserved genes present in genomic DNA using the methods described by Southern, J. Mol. Biol., 98:503 (1975). If the gene used as a hybridization probe hybridizes

to multiple restriction endonuclease fragments of the genome, that gene is a member of a repertoire of conserved genes.

Immunoglobulins

The immunoglobulins, or antibody molecules, are a large family of molecules that include several types of molecules, such as IgD, IgG, IgA, IgM and IgE. The antibody molecule is typically comprised of two heavy (H) and light (L) chains with both a variable (V) and constant (C) region present on each chain as shown in Figure 1. Schematic diagrams of human IgG heavy chain and human kappa light chain are shown in Figures 2A and 2B, respectively. Several different regions of an immunoglobulin contain conserved sequences useful for isolating an immunoglobulin repertoire. Extensive amino acid and nucleic acid sequence data displaying exemplary conserved sequences is compiled for immunoglobulin molecules by Kabat et al., in Sequences of Proteins of Immunological Interest, National Institutes of Health, Bethesda, MD, 1987.

The C region of the H chain defines the particular immunoglobulin type. Therefore the selection of conserved sequences as defined herein from the C region of the H chain results in the preparation of a repertoire of immunoglobulin genes having members of the immunoglobulin type of the selected C region.

The V region of the H or L chain typically comprises four framework (FR) regions each containing relatively lower degrees of variability that includes lengths of conserved sequences. The use of conserved sequences from the FR1 and FR4 (J region) framework regions of the V_H chain is a preferred exemplary

embodiment and is described herein in the Examples. Framework regions are typically conserved across several or all immunoglobulin types and thus conserved sequences contained therein are particularly suited for preparing repertoires having several immunoglobulin types.

Major Histocompatibility Complex

The major histocompatibility complex (MHC) is a large genetic locus that encodes an extensive family of proteins that include several classes of molecules referred to as class I, class II or class III MHC molecules. Paul et al., in Fundamental Immunology, Raven Press, NY, pp. 303-378 (1984).

Class I MHC molecules are a polymorphic group of transplantation antigens representing a conserved family in which the antigen is comprised of a heavy chain and a non-MHC encoded light chain. The heavy chain includes several regions, termed the N, C1, C2, membrane and cytoplasmic regions. Conserved sequences useful in the present invention are found primarily in the N, C1 and C2 regions and are identified as continuous sequences of "invariant residues" in Kabat et al., supra.

Class II MHC molecules comprise a conserved family of polymorphic antigens that participate in immune responsiveness and are comprised of an alpha and a beta chain. The genes coding for the alpha and beta chain each include several regions that contain conserved sequences suitable for producing MHC class II alpha or beta chain repertoires. Exemplary conserved nucleotide sequences include those coding for amino acid residues 26-30 of the A1 region, residues 161-170 of the A2 region and residues 195-206 of the membrane region, all of the alpha chain.

Conserved sequences are also present in the B1, B2 and membrane regions of the beta chain at nucleotide sequences coding for amino acid residues 41-45, 150-162 and 200-209, respectively.

Lymphocyte Receptors and Cell Surface Antigens

Lymphocytes contain several families of proteins on their cell surfaces including the T-cell receptor, Thy-1 antigen and numerous T-cell surface antigens including the antigens defined by the monoclonal antibodies OKT4 (leu3), OKUT5/8 (leu2), OKUT3, OKUT1 (leu1), OKT 11 (leu5) OKT6 and OKT9. Paul, supra at pp. 458-479.

The T-cell receptor is a term used for a family of antigen binding molecules found on the surface of T-cells. The T-cell receptor as a family exhibits polymorphic binding specificity similar to immunoglobulins in its diversity. The mature T-cell receptor is comprised of alpha and beta chains each having a variable (V) and constant (C) region. The similarities that the T-cell receptor has to immunoglobulins in genetic organization and function shows that T-cell receptor contains regions of conserved sequence. Lai et al., Nature, 331:543-546 (1988).

Exemplary conserved sequences include those coding for amino acid residues 84-90 of alpha chain, amino acid residues 107-115 of beta chain, and amino acid residues 91-95 and 111-116 of the gamma chain. Kabat et al., supra, p. 279.

Integrins And Adhesions

Adhesive proteins involved in cell attachment are members of a large family of related proteins termed integrins. Integrins are heterodimers

comprised of a beta and an alpha subunit. Members of the integrin family include the cell surface glycoproteins platelet receptor GpIIb-IIIa, vitronectin, receptor (VnR) fibronectin receptor (FnR) and the leukocyte adhesion receptors LFA-1, Mac-1, Mo-1 and 60.3. Rouslahti et al., Science, 238:491-497 (1987). Nucleic acid and protein sequence data demonstrates regions of conserved sequences exist in the members of these families, particularly between the beta chain of GpIIb-IIIa VnR and FnR, and between the alpha subunit of VnR, Mac-1, LFA-1, FnR and GpIIb-IIIa. Suzuki et al., Proc. Natl. Acad. Sci. USA, 83:8614-8618, 1986; Ginsberg et al., J. Biol. Chem., 262:5437-5440, 1987.

Fusion PCR

In the present invention, fusion PCR is used to generate two PCR-amplified DNA fragments, each of which have one of their ends modified by directed mispriming so that those ends share regions of complementarity, i.e., cohesive termini. When the two fragments are mixed, denatured and reannealed in a PCR cycle, the cohesive termini on two strands hybridize to form an "overlapping" DNA duplex that is internally primed. The subsequent PCR cycle primer-extends the non-overlapping regions to form a hybrid DNA molecule that is dicistronic. See Figure 4.

PCR amplification methods are described in detail in U.S. Patent Nos. 4,683,192, 4,683,202, 4,800,159, and 4,965,188, and at least in several texts including "PCR Technology: Principles and Applications for DNA Amplification", H. Erlich, ed., Stockton Press, New York (1989); and "PCR Protocols: A Guide to Methods and Applications", Innis et al., eds., Academic Press, San Diego, California (1990).

Cloning From Gene Repertoires

The following discussion illustrates the method of the present invention applied to isolating a pair of V_H and V_L genes from the immunoglobulin gene repertoire. This discussion is not to be taken as limiting, but rather as illustrating application of principles that can be used to operatively link and isolate a functionally similar pair of genes. The illustrated method can be used with any family of conserved genes coding for functionally related dimeric receptors, whether obtained directly from a natural source, such naive or in vivo immunized cells, or from cells or one or more genes that have been treated or mutagenized in vitro. Generally, the method, combines the following elements:

1. Producing V_H and V_L gene repertoires.

2. Preparing sets of outside and inside polynucleotide primers for cloning polynucleotide segments containing immunoglobulin V_H and V_L region genes.

3. Preparing a library containing a plurality of different dicistronic DNA molecules, each containing a V_H and a V_L gene from the respective repertoires.

4. Expressing the dicistronic DNA molecules in suitable host cells.

5. Screening the receptors formed by the polypeptides expressed by the dicistronic DNA molecules for the preselected activity, and segregating a dicistronic DNA molecule identified by the screening process.

In one method of producing a library of dicistronic DNA molecules containing upstream and downstream cistrons, first and second PCR

amplification products are produced using respective first and second PCR primer pairs. The first PCR primer pair comprises a first polypeptide outside primer and a first polypeptide inside primer. Similarly, the second PCR primer pair comprises a second polypeptide outside primer and a second polypeptide inside primer. The first and second polypeptide inside primers contain complementary 5'-terminal sequences that allow their DNA complements to hybridize and form an internally-primed duplex having 3'-overhanging termini. The internally-primed duplex is then subjected to primer extension reaction conditions to produce a double stranded, dicistronic DNA having substantially blunt or blunt ends. The dicistronic DNA is then PCR amplified using the outside primers as a PCR primer pair.

A dicistronic DNA molecule of this invention contains two amino acid residue-coding sequences on the same strand separated by at least one stop codon and at least one signal sequence necessary for translation of the downstream cistron, such as a translation initiation codon, ribosome binding site, and the like. Thus, the upstream and downstream cistrons of the dicistronic DNA molecule are operatively linked by a cistronic bridge. The cistronic bridge contains the genetic elements necessary to terminate translation of the upstream cistron and initiate translation of the downstream cistron. For instance, the coding strand of the bridge codes for one or more stop codons, preferably two, in the same translational reading frame as the upstream cistron. The cistronic bridge coding strand preferably also encodes a ribosome binding site for the downstream cistron located downstream from the upstream cistron's stop codon(s). Typically, the

coding strand of the cistronic bridge will also encode a leader polypeptide segment in the same translational reading frame as the downstream cistron. When present, the nucleotide base sequence encoding the leader usually begins with an initiation codon located within an operative distance, i.e., is operatively linked, to the ribosome binding site.

A receptor produced by the present invention assumes a conformation having a binding site specific for, as evidenced by its ability to be competitively inhibited, a preselected or predetermined ligand such as an antigen, enzymatic substrate and the like. In one embodiment, a receptor of this invention is a ligand binding heterodimeric polypeptide that forms an antigen binding site which specifically binds to a preselected antigen to form a complex having a sufficiently strong binding between the antigen and the binding site for the complex to be isolated. When the receptor is an antigen binding polypeptide its affinity or avidity is generally greater than 10^5 M^{-1} more usually greater than 10^6 M^{-1} and preferably greater than 10^8 M^{-1}.

In another embodiment, a receptor of the subject invention binds a substrate and catalyzes the formation of a product from the substrate. While the topology of the ligand binding site of a catalytic receptor is probably more important for its preselected activity than its affinity (association constant or pKa) for the substrate, the subject catalytic receptors have an association constant for the preselected substrate generally greater than 10^3 M^{-1}, more usually greater than 10^5 M^{-1} or 10^6 M^{-1} and preferably greater than 10^7 M^{-1}.

Preferably the receptor produced by the subject invention is heterodimeric and is therefore

normally comprised of two different polypeptide chains, which together assume a conformation having a binding affinity, or association constant for the preselected ligand that is different, preferably higher, than the affinity or association constant of either of the polypeptides alone, i.e., as monomers. One or both of the different polypeptide chains is derived from the variable region of the light and heavy chains of an immunoglobulin. Typically, polypeptides comprising the light (V_L) and heavy (V_H) variable regions are employed together for binding the preselected ligand.

A receptor produced by the subject invention can be comprised of active monomers V_H and V_L ligand binding polypeptides produced by the present invention can be advantageously combined in the heterodimer to modulate the activity of either or to produce an activity unique to the heterodimer.

The individual ligand polypeptides will be referred to as V_H and V_L and the heterodimer will be referred to as a F_V. However, it should be understood that a V_H may contain in addition to the V_H, substantially all or a portion of the heavy chain constant region. Similarly, a V_L may contain, in addition to the V_L, substantially all or a portion of the light chain constant region. A heterodimer comprised of a V_H containing a portion of the heavy chain constant region and a V_L containing substantially all of the light chain constant region is termed a Fab fragment. The production of Fab can be advantageous in some situations because the additional constant region sequences contained in a Fab as compared to a F_V can stabilize the V_H and V_L interaction. Such stabilization can cause the Fab to have higher affinity for antigen. In addition the Fab

is more commonly used in the art and thus there are more commercial antibodies available to specifically recognize a Fab in screening procedures.

The individual V_H and V_L polypeptides can be produced in lengths equal to or substantially equal to their naturally occurring lengths. See Figure 2. However, in preferred embodiments, the V_H and V_L polypeptides will generally have fewer than 125 amino acid residues, more usually fewer than about 120 amino acid residues, while normally having greater than 60 amino acid residues, usually greater than about 95 amino acid residues, more usually greater than about 100 amino acid residues. Preferably, the V_H will be from about 110 to about 125 amino acid residues in length while V_L will be from about 95 to about 115 amino acid residues in length.

The amino acid residue sequences will vary widely, depending upon the particular idiotype involved. Usually, there will be at least two cysteines separated by from about 60 to 75 amino acid residues and joined by a disulfide bond. The polypeptides produced by the subject invention will normally be substantial copies of idiotypes of the variable regions of the heavy and/or light chains of immunoglobulins, but in some situations a polypeptide may contain random mutations in amino acid residue sequences in order to advantageously improve the desired activity.

In some situations, it is desirable to provide for covalent cross linking of the V_H and V_L polypeptides, which can be accomplished by providing cysteine resides at the carboxyl termini. The polypeptide will normally be prepared free of the immunoglobulin constant regions, however a small portion of the J region may be included as a result of

the advantageous selection of DNA synthesis primers. The D region will normally be included in the transcript of the V_H.

Typically the C terminus region of the V_H and V_L polypeptides will have a greater variety of sequences than the N terminus and, based on the present strategy, can be further modified to permit a variation of the normally occurring V_H and V_L chains. A synthetic polynucleotide can be employed to vary one or more amino acid in a hypervariable region.

1. Producing A Gene Repertoire

A gene repertoire useful in practicing the present invention contains at least 10^3, preferably at least 10^4, more preferably at least 10^5, and most preferably at least 10^7 different conserved genes. Methods for evaluating the diversity of a repertoire of conserved genes is well known to one skilled in the art.

Various well known methods can be employed to produce a useful gene repertoire. For instance, V_H and V_L gene repertoires can be produced by isolating V_H- and V_L-coding mRNA from a heterogeneous population of antibody producing cells, i.e., B lymphocytes (B cells), preferably rearranged B cells such as those found in the circulation or spleen of a vertebrate. Rearranged B cells are those in which immunoglobulin gene translocation, i.e., rearrangement, has occurred as evidenced by the presence in the cell of mRNA with the immunoglobulin gene V, D and J region transcripts adjacently located thereon. Typically, the B cells are collected in a 1-100 ml sample of blood which usually contains 10^6 B cells/ml.

In some cases, it is desirable to bias a repertoire for a preselected activity, such as by

using as a source of nucleic acid cells (source cells) from vertebrates in any one of various stages of age, health and immune response. For example, repeated immunization of a healthy animal prior to collecting rearranged B cells results in obtaining a repertoire enriched for genetic material producing a receptor of high affinity. Mullinax et al., Proc. Natl. Acad. Sci. USA, 87:8095-8099 (1990). Conversely, collecting rearranged B cells from a healthy animal whose immune system has not been recently challenged results in producing a repertoire that is not biased towards the production of high affinity V_H and/or V_L polypeptides.

It should be noted the greater the genetic heterogeneity of the population of cells for which the nucleic acids are obtained, the greater the diversity of the immunological repertoire (comprising V_H- and V_L-coding genes) that will be made available for screening according to the method of the present invention. Thus, cells from different individuals, particularly those having an immunologically significant age difference, and cells from individuals of different strains, races or species can be advantageously combined to increase the heterogeneity (diversity) of a repertoire.

Thus, in one preferred embodiment, the source cells are obtained from a vertebrate, preferably a mammal, which has been immunized or partially immunized with an antigenic ligand (antigen) against which activity is sought, i.e., a preselected antigen. The immunization can be carried out conventionally. Antibody titer in the animal can be monitored to determine the stage of immunization desired, which stage corresponds to the amount of enrichment or biasing of the repertoire desired. Partially immunized animals typically receive only one

immunization and cells are collected from those animals shortly after a response is detected. Fully immunized animals display a peak titer, which is achieved with one or more repeated injections of the antigen into the host mammal, normally at 2 to 3 week intervals. Usually three to five days after the last challenge, the spleen is removed and the genetic repertoire of the spleenocytes, about 90% of which are rearranged B cells, is isolated using standard procedures. See, Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, NY. Nucleic acids coding for V_H and V_L polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, most preferably from IgM and IgG, producing cells.

Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned as a diverse population are well known in the art. See for example Herrmann et al., Methods In Enzymol., 152:180-183, (1987); Frischauf, Methods In Enzymol., 152:183-190 (1987); Frischauf, Methods In Enzymol., 152:190-199 (1987); and DiLella et al., Methods In Enzymol., 152:199-212 (1987). (The teachings of the references cited herein are hereby incorporated by reference.)

The desired gene repertoire can be isolated from either genomic material containing the gene expressing the variable region or the messenger RNA (mRNA) which represents a transcript of the variable region. The difficulty in using the genomic DNA from other than non-rearranged B lymphocytes is in juxtaposing the sequences coding for the variable region, where the sequences are separated by introns. The DNA fragment(s) containing the proper exons must be isolated, the introns excised, and the exons then

spliced in the proper order and in the proper orientation. For the most part, this will be difficult, so that the alternative technique employing rearranged B cells will be the method of choice because the V, D and J immunoglobulin gene regions have translocated to become adjacent, so that the sequence is continuous (free of introns) for the entire variable regions.

Where mRNA is utilized the cells will be lysed under RNase inhibiting conditions. In one embodiment, the first step is to isolate the total cellular mRNA. Poly A+ mRNA can then be selected by hybridization to an oligo-dT cellulose column. The presence of mRNAs coding for the heavy and/or light chain polypeptides can then be assayed by hybridization with DNA single strands of the appropriate genes. Conveniently, the sequences coding for the constant portion of the V_H and V_L can be used as polynucleotide probes, which sequences can be obtained from available sources. See for example, Early and Hood, Genetic Engineering, Setlow and Hollaender, eds., Vol. 3, Plenum Publishing Corporation, NY, (1981), pages 157-188; and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, MD, (1987). In preferred embodiments, the preparation containing the total cellular mRNA is first enriched for the presence of V_H and/or V_L coding mRNA. Enrichment is typically accomplished by subjecting the total mRNA preparation or partially purified mRNA product thereof to a primer extension reaction employing a polynucleotide synthesis primer of the present invention. Exemplary methods for producing V_H and V_L gene repertoires are described in PCT Application No. PCT/US 90/02836 (International Publication No. WO 90/14430).

In preferred embodiments, isolated B cells are immunized in vitro against a preselected antigen. In vitro immunization is defined as the clonal expansion of epitope-specific B cells in culture, in response to antigen stimulation. The end result is to increase the frequency of antigen-specific B cells in the immunoglobulin repertoire, and thereby decrease the number of clones in an expression library that must be screened to identify a clone expressing an antibody of the desired specificity. The advantage of in vitro immunization is that human monoclonal antibodies can be generated against a limitless number of therapeutically valuable antigens, including toxic or weak immunogens. For example, antibodies specific for the polymorphic determinants of tumor-associated antigens, rheumatoid factors, and histocompatibility antigens can be produced, which can not be elicited in immunized animals. In addition, it may be possible to generate immune responses which are normally suppressed in vivo.

In vitro immunization can be used to give rise to either a primary or secondary immune response. A primary immune response, resulting from first time exposure of a B cell to an antigen, results in clonal expansion of epitope-specific cells and the secretion of IgM antibodies with low to moderate apparent affinity constants (10^6-$10^8 M^{-1}$). Primary immunization of human splenic and tonsillar lymphocytes in culture can be used to produce monoclonal antibodies against a variety of antigens, including cells, peptides, macromolecules, haptens, and tumor-associated antigens. Memory B cells from immunized donors can also be stimulated in culture to give rise to a secondary immune response characterized by clonal expansion and the production of high affinity

antibodies (>10^9 M^{-1}) of the IgG isotype, particularly against viral antigens by clonally expanding sensitized lymphocytes derived from seropositive individuals.

In one embodiment, peripheral blood lymphocytes are depleted of various cytolytic cells that appear to down-modulate antigen-specific B cell activation. When lysosome-rich subpopulations (natural killer cells, cytotoxic and suppressor T cells, monocytes) are first removed by treatment with the lysosmotropic methyl ester of leucine, the remaining cells (including B cells, T helper cells, accessory cells) respond antigen-specifically during in vitro immunization. The lymphokine requirements for inducing antibody production in culture are satisfied by a culture supernatant from activated, irradiated T cells.

In addition to in vitro immunization, cell panning (immunoaffinity absorption) can be used to further increase the frequency of antigen-specific B cells. Techniques for selecting B cell subpopulations via solid-phase antigen binding are well established. Panning conditions can be optimized to selectively enrich for B cells which bind with high affinity to a variety of antigens, including cell surface proteins. Panning can be used alone, or in combination with in vitro immunization to increase the frequency of antigen-specific cells above the levels which can be obtained with either technique alone. Immunoglobulin expression libraries constructed from enriched populations of B cells are biased in favor of antigen-specific antibody clones, and thus, enabling identification of clones with the desired specificities from smaller, less complex libraries.

2. Preparation Of Polynucleotide Primers

The term "polynucleotide" as used herein in reference to primers, probes and nucleic acid fragments or segments to be synthesized by primer extension is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than 3. Its exact size will depend on many factors, which in turn depends on the ultimate conditions of use.

The term "primer" as used herein refers to a polynucleotide whether purified from a nucleic acid restriction digest or produced synthetically, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase, reverse transcriptase and the like, and at a suitable temperature and pH. The primer is preferably single stranded for maximum efficiency, but may alternatively be in double stranded form. If double stranded, the primer is first treated to separate it from its complementary strand before being used to prepare extension products. Preferably, the primer is a polydeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agents for polymerization. The exact lengths of the primers will depend on may factors, including temperature and the source of primer. For example, depending on the complexity of the target sequence, a polynucleotide primer typically contains 15 to 25 or more nucleotides, although it can contain fewer nucleotides. Short primer molecules generally require

cooler temperatures to form sufficiently stable hybrid complexes with template.

The primers used herein are selected to be "substantially" complementary to the different strands of each specific sequence to be synthesized or amplified. This means that the primer must be sufficiently complementary to non-randomly hybridize with its respective template strand. Therefore, the primer sequence may or may not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment can be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. Such non-complementary fragments typically code for an endonuclease restriction site. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided the primer sequence has sufficient complementarily with the sequence of the strand to be synthesized or amplified to non-randomly hybridize therewith and thereby form an extension product under polynucleotide synthesizing conditions.

Primers of the present invention may also contain a DNA-dependent RNA polymerase promoter sequence or its complement. See for example, Krieg et al., Nucleic Acids Research, 12:7057-70 (1984); Studier et al., J. Mol. Biol., 189:113-130 (1986); and Molecular Cloning: A Laboratory Manual, Second Edition, Maniatis et al., eds., Cold Spring Harbor, NY (1989).

When a primer containing a DNA-dependent RNA polymerase promoter is used the primer is hybridized to the polynucleotide strand to be amplified and the second polynucleotide strand of the DNA-dependent RNA polymerase promoter is completed using an inducing

agent such as E. coli DNA polymerase I, or the Klenow fragment of E. coli DNA polymerase. The starting polynucleotide is amplified by alternating between the production of an RNA polynucleotide and DNA polynucleotide.

Primers may also contain a template sequence or replication initiation site for a RNA-directed RNA polymerase. Typical RNA-directed RNA polymerase include the QB replicase described by Lizardi et al., Biotechnology, 6:1197-1202 (1988). RNA-directed polymerases produce large numbers of RNA strands from a small number of template RNA strands that contain a template sequence or replication initiation site. These polymerases typically give a one million-fold amplification of the template strand as has been described by Kramer et al., J. Mol. Biol., 89:719-736 (1974).

The polynucleotide primers can be prepared using any suitable method, such as, for example, the phosphotriester or phosphodiester methods see Narang et al., Meth. Enzymol., 68:90, (1979); U.S. Patent No. 4,356,270; and Brown et al., Meth. Enzymol., 68:109, (1979).

The choice of a primer's nucleotide sequence depends on factors such as the distance on the nucleic acid from the region coding for the desired receptor, its hybridization site on the nucleic acid relative to any second primer to be used, the number of genes in the repertoire it is to hybridize to, and the like.

(a) Primers for Producing Gene Repertoires

V_H and V_L gene repertoires can be separately prepared prior to their utilization in the present invention. Repertoire preparation is typically accomplished by primer extension, preferably

by primer extension in a PCR format.

To produce a repertoire of V_H-coding DNA homologs by primer extension, the nucleotide sequence of a primer is selected to hybridize with a plurality of immunoglobulin heavy chain genes at a site substantially adjacent to the V_H-coding region so that a nucleotide sequence coding for a functional (capable of binding) polypeptide is obtained. To hybridize to a plurality of different V_H-coding nucleic acid strands, the primer must be a substantial complement of a nucleotide sequence conserved among the different strands. Such sites include nucleotide sequences in the constant region, any of the variable region framework regions, preferably the third framework region, leader region, promoter region, J region and the like.

If the repertoires V_H-coding and V_L-coding DNA homologs are to be produced by polymerase chain reaction (PCR) amplification, two primers, i.e., a PCR primer pair, must be used for each coding strand of nucleic acid to be amplified. The first primer becomes part of the nonsense (minus or complementary) strand and hybridizes to a nucleotide sequence conserved among V_H (plus or coding) strands within the repertoire. To produce V_H coding DNA homologs, first primers are therefore chosen to hybridize to (i.e. be complementary to) conserved regions within the J region, CH1 region, hinge region, CH2 region, or CH3 region of immunoglobulin genes and the like. To produce a V_L coding DNA homolog, first primers are chosen to hybridize with (i.e. be complementary to) a conserved region within the J region or constant region of immunoglobulin light chain genes and the like. Second primers become part of the coding (plus) strand and hybridize to a nucleotide sequence

conserved among minus strands. To produce the V_H-coding DNA homologs, second primers are therefore chosen to hybridize with a conserved nucleotide sequence at the 5' end of the V_H-coding immunoglobulin gene such as in that area coding for the leader or first framework region. It should be noted that in the amplification of both V_H- and V_L-coding DNA homologs the conserved 5' nucleotide sequence of the second primer can be complementary to a sequence exogenously added using terminal deoxynucleotidyl transferase as described by Loh et al., Sci. Vol 243:217-220 (1989). One or both of the first and second primers can contain a nucleotide sequence defining an endonuclease recognition site. The site can be heterologous to the immunoglobulin gene being amplified and typically appears at or near the 5' end of the primer.

(b) Inside and Outside Primers

In one embodiment, the present invention utilizes a set of polynucleotides that form inside primers comprised of an upstream inside primer and a downstream inside primer. Each of the inside primers has a priming region located at the 3'-terminus of the primer. The priming region is typically the 3'-most (3'-terminal) 15 to 30 nucleotide bases. The 3'-terminal priming portion of each inside primer is capable of acting as a primer to catalyze nucleic acid synthesis, i.e., initiate a primer extension reaction off its 3' terminus. One or both of the inside primers is further characterized by the presence of a 5'-terminal (5'-most) non-priming portion, i.e., a region that does not participate in hybridization to repertoire template.

In fusion PCR, each inside primer works in combination with an outside primer to amplify a target nucleic acid sequence. The choice of PCR primer pairs for use in fusion PCR as described herein is governed by the same considerations as previously discussed for choosing PCR primer pairs useful in producing gene repertoires. That is, the primers have a nucleotide sequence that is complementary to a sequence conserved in the repertoire. Useful V_L and V_H inside priming sequences are shown in Tables 1 and 2, respectively, below.

Table 1

3' Priming Portions of Various Inside V_L Primers

Seq. Id. No.	
(1)[1]	5' GTGATGACCCACTCTCC 3'
(2)	5' GTGATGACCCAGTCTCCA 3'
(3)	5' GTTGTGACTCAGGAATCT 3'
(4)	5' GTGTTGACGCAGCCGCCC 3'
(5)	5' GTGCTCACCCAGTCTCCA 3'
(6)	5' CAGATGACCCAGTCTCCA 3'
(7)	5' GTGATGACCCAGACTCCA 3'
(8)	5' GTCATGACCCAGTCTCCA 3'
(9)	5' TTGATGACCCAAACTCAA 3'
(10)	5' GTGATAACCCAGGATGAA 3'

[1] Nucleotide sequences 1-10 are unique 5' primers for the amplification of kappa light chain variable regions.

Table 2

3' Priming Portions of Various Inside V_H Primers

	Seq. Id. No.	
	(11)[1]	5' ACAAGATTTGGGCTC 3'
	(12)[2]	5' TGGGGTTTTGAGCTC 3'
	(13)[3]	5' GAGACAGTGACCGGGTTCCTTGGCCCCA 3
	(14)[4]	5' TGGAATGGGCACATGCAG 3'
	(15)[5]	5' TTATCATTTACCCGGAGA 3'
	(16)[6]	5' AACGGTAACAGTGGTGCCTTGGCCCCA 3'
	(17)[7]	5' ACAATCCCTGGGCACAAT 3'
	(18)[8]	5' CACCTTGGTGCTGCTGGC 3'
	(19)[9]	5' ACAACCACAATCCCTGGGCACAATTTT 3'
	(20)[10]	5' ACAATCCCTGGGCACAAT 3'
	(21)[11]	5' GAGTTCACTAGTTGGGCACGGTGGGCA 3'

[1] Unique 3' primer for human IgG1, 2, 3 and 4 Fd.

[2] Unique 3' primer for human V_H amplification.

[3] 3' primer for amplifying human heavy chain variable regions.

[4] 3' primer for amplifying the Fd region of mouse IgM.

[5] 3' primer located in the CH3 region of human IgG1 to amplify the entire heavy chain.

[6] Unique 3' primer for amplification of mouse F_V.

[7] Unique 3' primer for amplification of mouse IgG1 Fd.

[8] Unique 3' primer for amplification of V_H including part of the mouse gamma 1 first constant region.

[9] Unique 3' primer for amplification of V_H including part of mouse gamma 1 first constant region and hinge region.

[10] 3' primer for amplifying mouse Fd including part of the mouse IgG first constant region and part of the hinge region.

11 3' primer for amplifying human IgG1 Fd including part of the human IgG first constant region and part of the hinge region including the two cysteines which create the disulfide bridge for producing Fab'2 (the primer corresponds to Kabat numbers 241QQ to 247).

A preferred set of inside primers used herein has primers with complementary 5'-terminal non-priming regions, the complementary strands of which are capable of hybridizing to each other to form a duplex with 3' overhangs. The duplex encodes all or part of a double stranded cistronic bridge. That is, if the 3' overhangs of the duplex are filled in with complementary bases so as to define a double stranded DNA extending from the 3'-terminus of one of the inside primers to the 3'-terminus of the other of the inside primers, that double stranded DNA segment forms a sequence of nucleotides that operatively links the upstream and downstream cistrons for polycistronic expression. Thus, while each of the inside primers in a set contains only a portion of the sequence information necessary to form the double stranded cistronic bridge, the two inside primers in combination encode both the plus and minus strands of all or part of the bridge.

For example, one inside upstream primer can have a sequence that forms a portion of the plus strand of the bridge, and the other inside primer encodes the sequence, through complementarity, of the downstream portion of the plus strand.

In a preferred embodiment, the plus strand of the cistronic bridge contains, in the translational reading frame and from an upstream position to a downstream position, sequences coding for (i) at least one stop codon, preferably two, in the same reading frame as the upstream cistron, (ii) a ribosome binding

site, and (iii) a polypeptide leader, the translation initiation codon of which is in the same reading frame as the downstream cistron. The stop codon is present to terminate translation of the upstream cistron. The ribosome binding site is present to initiate translation of the downstream cistron from the polycistronic mRNA.

The predicted amino acid residue sequences of two pelB gene product variants from Erwinia Carotova are shown in Table 3. Lei, et al., supra., Amino Acid residue sequences for other leaders from E. coli useful in this invention are also listed in Table 3. Oliver, In Neidhart, F. C. (ed.), Escherichia coli and Salmonella Typhimurium, American Society for Microbiology, Washington, D.C., 1:56-69 (1987). These regions for the heavy chain are contained in the modified ImmunoZAP H expression vector. Mullinax, et al., Proc. Natl. Acad. Sci., USA. 87:8095-8099 (1990).

Table 3

Leader Sequences

Seq. Id. No.	Type	Amino Acid Residue Sequence
(22)	pelB[1]	MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeu LeuLeuAlaAlaGlnProAlaGlnProAlaMetAla
(23)	pelB[2]	MetLysSerLeuIleThrProIleAlaAlaGlyLeuLeu LeuAlaPheSerGlnTyrSerLeuAla
(24)	MalE[3]	MetLysIleLysThrGlyAlaArgIleLeuAlaLeuSer AlaLeuThrThrMetMetPheSerAlaSerAlaLeuAla LysIle
(25)	OmpF[3]	MetMetLysArgAsnIleLeuAlaValIleValProAla LeuLeuValAlaGlyThrAlaAsnAlaAlaGlu
(26)	PhoA[3]	MetLysGlnSerThrIleAlaLeuAlaLeuLeuProLeu LeuPheThrProValThrLysAlaArgThr

(27)	Bla^3	MetSerIleGlnHisPheArgValAlaLeuIleProPhe PheAlaAlaPheCysLeuProValPheAlaHisPro
(28)	$LamB^3$	MetMetIleThrLeuArgLysLeuProLeuAlaValAla ValAlaAlaGlyValMetSerAlaGlnAlaMetAlaVal Asp
(29)	Lpp^3	MetLysAlaThrLysLeuValLeuGlyAlaValIleLeu GlySerThrLeuLeuAlaGlyCysSer

[1] pelB from Erwinia carotovora gene

[2] pelB from Erwinia carotovora EC 16 gene

[3] leader sequences from E. coli

To achieve high levels of gene expression in E. coli, it is necessary to use not only strong promoters to generate large quantities of mRNA, but also ribosome binding sites to ensure that the mRNA is efficiently translated. In E. coli, the ribosome binding site includes an initiation codon (AUG) and a sequence 3-9 nucleotides long located 3 11 nucleotides upstream from the initiation codon [Shine et al., Nature, 254:34 (1975). The sequence, AGGAGGU, which is called the Shine-Dalgarno (SD) sequence, is complementary to the 3' end of E. coli 16S mRNA. Binding of the ribosome to mRNA and the sequence at the 3' end of the mRNA can be affected by several factors:

(i) The degree of complementarity between the SD sequence and 3' end of the 16S tRNA.

(ii) The spacing and possibly the DNA sequence lying between the SD sequence and the AUG [Roberts et al., Proc. Natl. Acad. Sci. USA, 76:760 (1979a); Roberts et al., Proc. Natl. Acad. Sci. USA, 76:5596 (1979b); Guarente et al., Science, 209:1428 (1980); and Guarente et al., Cell, 20:543 (1980).] Optimization is achieved by measuring the level of

expression of genes in plasmids in which this spacing is systematically altered. Comparison of different mRNAs shows that there are statistically preferred sequences from positions -20 to +13 (where the A of the AUG is position 0) [Gold et al., Annu. Rev. Microbiol., 35:365 (1981)]. Leader sequences have been shown to influence translation dramatically (Roberts et al., 1979 a, b supra).

(iii) The nucleotide sequence following the AUG, which affects ribosome binding [Taniguchi et al., J. Mol. Biol., 118:533 (1978)].

Useful ribosome binding sites are shown in Table 4 below.

Table 4

Ribosome Binding Sites*

	Seq. Id. No.		
1.	(30)	5'	AAUCUUGGAGGCUUUUUUAUGGUUCGUUCU
2.	(31)	5'	UAACUAAGGAUGAAAUGCAUGUCUAAGACA
3.	(32)	5'	UCCUAGGAGGUUUGACCUAUGCGAGCUUUU
4.	(33)	5'	AUGUACUAAGGAGGUUGUAUGGAACAACGC

* Sequences of initiation regions for protein synthesis in four phage mRNA molecules are underlined.

AUG = initiation codon (double underlined)

1. = Phage ϕX174 gene-A protein
2. = Phage Qβ replicase
3. = Phage R17 gene-A protein
4. = Phage lambda gene-cro protein

It is preferred that the complementary (overlapping) region of the inside primers and the priming portion of the inside primers have about the same denaturation temperature, Td. The Td of a

sequence can be estimated by the following formula: Td = 4(C+G) + 2(A+T), where C, G, A and T represent the respective number of cytosine, guanine, adenine and thymine bases in the sequence. A Td for the above-identified hybridizing region of about 45-55°C, preferably about 50°C, is preferred. Typically, overlapping regions in the range of about 15 to 20 nucleotides works well in conjunction with priming regions in the range of 15-30 nucleotides.

The set of outside primers forms the termini of the dicistronic DNA molecule. The set of outside primers comprises an upstream outside primer and a downstream outside primer. The outside primers each comprise a 3'-terminal priming portion, and preferably a portion that defines an endonuclease restriction site. When present, the restriction site-defining portion is typically located in a 5'-terminal non-priming portion of the outside primer. The restriction site defined by the upstream outside primer is typically chosen to be one recognized by a restriction enzyme that does not recognize the restriction site defined by the downstream outside primer, the objective being to be able to produce a dicistronic DNA having cohesive termini that are non-complementary to each other and thus allow directional insertion into a vector.

Useful outside primer sequences are shown in Tables 5 and 6 below.

Table 5

Outside V_H Primers

Seq. Id. No.	
(34)[1]	5'AGGTCCAGCTGCTCGAGTCTGG3'
(35)	5'AGGTCCAGCTGCTCGAGTCAGG3'
(36)	5'AGGTCCAGCTTCTCGAGTCTGG3'
(37)	5'AGGTCCAGCTTCTCGAGTCAGG3'
(38)	5'AGGTCCAACTGCTCGAGTCTGG3'
(39)	5'AGGTCCAACTGCTCGAGTCAGG3'
(40)	5'AGGTCCAACTTCTCGAGTCTGG3'
(41)	5'AGGTCCAACTTCTCGAGTCAGG3'
(42)[2]	5'AGGTGCAGCTGCTCGAGTCTGG3'
(43)	5'AGGTGCAGCTGCTCGAGTCGGG3'
(44)	5'AGGTGCAACTGCTCGAGTCTGG3'
(45)	5'AGGTGCAACTGCTCGAGTCGGG3'

[1] Nucleotide sequences 21-28 are unique 5' primers for the amplification of mouse V_H genes.

[2] Nucleotide sequences 29-32 are unique 5' primers for amplification of nucleic acids coding for human variable regions.

Table 6

Outside V_L Primers

Seq. Id. No.	
(46)[1]	5' ACGTCTAGATTCCACCTTGGTCCC 3'
(47)[2]	5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'
(48)[3]	5' GCATTCTAGACTATTAACATTCTGTAGGGGC 3'
(49)[4]	5' GCAGCATTCTAGAGTTTCAGCTCCAGCTTGCC 3'
(50)[5]	5' CCGCCGTCTAGAACACTCATTCCTGTTGAAGCT 3'
(51)[6]	5' CCGCCGTCTAGAACATTCTGCAGGAGACAGACT 3'

(52)[7] 5' GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA 3'

(53)[8] 5' GCCGCTCTAGAACACTCATTCCTGTTGAA 3'

(54)[9] 5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA 3'

(55)[10] 5' GCATTCTAGACTATTATGAACATTCTGTAGGGGC 3'

[1] 3' primer for amplifying human kappa chain variable regions.

[2] 3' primer in human kappa light chain constant region.

[3] 3' primer in human lambda light chain constant region.

[4] Unique 3' primer for amplification of kappa light chain variable regions.

[5] Unique 3' primer for mouse kappa light chain amplification including the constant region.

[6] Unique 3' primer for mouse lambda light chain amplification including the constant region.

[7] Unique 3' primer for amplification of kappa light chain.

[8] Unique 3' primer for amplification of mouse kappa light chain.

[9] Unique 3' primer for kappa V_L amplification.

[10] Unique 3' primer for human, mouse and rabbit lambda V_L amplification.

3. Preparing a Dicistronic DNA Molecule Library

The strategy used for cloning the V_H and V_L genes contained within a repertoire will depend, as is well known in the art, on the type, complexity, and purity of the nucleic acids making up the repertoire. Other factors include whether or not the genes are contained in one or a plurality of repertoires and whether or not they are to be amplified and/or mutagenized.

In one embodiment, a library of dicistronic DNA molecules containing upstream and downstream cistrons operatively linked by a cistronic bridge can be produced by the following steps:

(a) Subjecting a repertoire of first polypeptide genes (e.g., V_H-coding genes), to PCR amplification using first outside and first inside primers, i.e., a first PCR primer pair, to form a first primary PCR product.

(b) Subjecting a repertoire of second polypeptide genes (e.g., V_L-coding genes) to PCR amplification using second outside and second inside primers, i.e., a second PCR primer pair, to form a second primary PCR product.

(c) Hybridizing the first and second primary PCR products to form internally (self) primed duplexes, i.e., duplexes having 3'-hybridized and 5'-overhanging termini.

(d) Subjecting the internally-primed duplexes to primer extension reaction conditions to form double stranded duplexes having substantially blunt, preferably blunt, termini and a dicistronic strand containing the upstream and downstream cistrons linked by a cistronic bridge encoded by the inside primers. By "substantially blunt" is meant having no more than about one or two overhanging nucleotides. (Substantially blunt double stranded DNA is sometimes produced by primer overextension by Taq polymerase, usually by the addition of one or two terminal adenine residues.)

The V_H- and V_L-coding gene repertoires are comprised of polynucleotide coding strands, such as mRNA and/or the sense strand of genomic DNA. If the repertoire is in the form of double stranded genomic DNA, it is usually first denatured, typically by melting, into single strands. A repertoire is

subjected to a PCR reaction by treating (contacting) the repertoire with a PCR primer pair, each member of the pair having a preselected nucleotide sequence. The PCR primer pair is capable of initiating primer extension reactions by hybridizing to nucleotide sequences, preferably at least about 10 nucleotides in length and more preferably at least about 20 nucleotides in length, conserved within the repertoire. The first primer of a PCR primer pair is sometimes referred to herein as the "sense primer" because it hybridizes to the coding or sense strand of a nucleic acid. In addition, the second primer of a PCR primer pair is sometimes referred to herein as the "anti-sense primer" because it hybridizes to a non-coding or anti-sense strand of a nucleic acid, i.e., a strand complementary to a coding strand.

The PCR reaction is performed by mixing the PCR primer pair, preferably a predetermined amount thereof, with the nucleic acids of the repertoire, preferably a predetermined amount thereof, in a PCR buffer to form a PCR reaction admixture. The admixture is maintained under polynucleotide synthesizing conditions for a time period, which is typically predetermined, sufficient for the formation of a PCR reaction product, thereby producing a plurality of different V_H-coding and/or V_L-coding DNA homologs.

A plurality of first primer and/or a plurality of second primers can be used in each amplification, e.g., one species of first primer can be paired with a number of different second primers to form several different primer pairs. Alternatively, an individual pair of first and second primers can be used. In any case, the amplification products of amplifications using the same or different combinations of first and

second primers can be combined to increase the diversity of the gene library.

In another strategy, the object is to clone the V_H- and/or V_L-coding genes from a repertoire by providing a polynucleotide complement of the repertoire, such as the anti-sense strand of genomic dsDNA or the polynucleotide produced by subjecting mRNA to a reverse transcriptase reaction. Methods for producing such complements are well known in the art.

The PCR reaction is performed using any suitable method. Generally it occurs in a buffered aqueous solution, i.e., a PCR buffer, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess (for genomic nucleic acid, usually about 10^6:1 primer:template) of the primer is admixed to the buffer containing the template strand. A large molar excess is preferred to improve the efficiency of the process.

In preferred embodiments the ratio of gene molecules and their respective primers is as follows: about 1 x 10^3 V_H gene molecules to about 1 x 10^8 outside V_H primer molecules, about 1 x 10^3 V_H gene molecules, to about 1 x 10^7 inside V_H gene primer molecules, about 1 x 10^3 V_L gene molecules to about 1 x 10^8 outside V_L gene primer molecules, about 1 x 10^4 V_L gene molecules to about 1 x 10^7 V_L gene primer molecules. In more preferred embodiments, 10^4 outside V_H gene primer molecules and 10^3 inside V_H gene primer molecules are used for every V_H gene molecule present in the PCR admixture. Similarly, 10^4 outside V_L gene primer molecules and 10^3 V_L inside gene primer molecules are used for every V_L gene molecule present in the PCR admixture. Thus, there is typically a 10 fold molar excess of outside primer to inside primer.

In the fusion PCR reaction, the gene repertoires are admixed with outside and inside primers, the outside primers being present in excess relative to the inside primers. The initial PCR thermocycles produce intermediate products having complementary termini from each of the first and second gene repertoires. That is, the end of one strand from one primary PCR product is capable of hybridizing with the complementary end from the other primary PCR product. The strands having the overlap at their 3' ends can act as primers for one another, i.e., from an internally primed duplex, and be extended by the polymerase to form the full length final product. The final product is then amplified by the set of outside primers, which act as a third PCR pair when the inside primers have been exhausted, to form a secondary PCR product. Typically the molar ratio of outside primers to inside primers is such that the inside primers are effectively exhausted within about 2 to about 12, preferably about 5, 6 or 7 thermocycles.

The PCR buffer also contains the deoxyribonucleotide triphosphates dATP, dCTP, dGTP, and dTTP and a polymerase, typically thermostable, all in adequate amounts for primer extension (polynucleotide synthesis) reaction. The resulting solution (PCR admixture) is heated to about 90°C - 100°C for about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to 54°C, which is preferable for primer hybridization. The synthesis reaction may occur at from room temperature up to a temperature above which the polymerase (inducing agent) no longer functions efficiently. Thus, for example, if DNA polymerase is used as inducing agent, the temperature

is generally no greater than about 40°C. An exemplary PCR buffer comprises the following: 50 mM KCl; 10 mM Tris-HCl; pH 8.3; 1.5 mM $MgCl_2$; 0.001% (wt/vol) gelatin, 200 μM dATP; 200 μM dTTP; 200 μM dCTP; 200 μM dGTP; and 2.5 units Thermus aquaticus DNA polymerase I (U.S. Patent No. 4,889,818) per 100 microliters of buffer.

The inducing agent may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be inducing agents, however, which initiate synthesis at the 5' end and proceed in the above direction, using the same process as described above.

The inducing agent also may be a compound or system which will function to accomplish the synthesis of RNA primer extension products, including enzymes. In preferred embodiments, the inducing agent may be a DNA-dependent RNA polymerase such as T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase. These polymerases produce a complementary RNA polynucleotide. The high turn over rate of the RNA polymerase amplifies the starting polynucleotide as has been described by Chamberlin et al., The Enzymes,

ed. P. Boyer, PP. 87-108, Academic Press, New York (1982). Another advantage of T7 RNA polymerase is that mutations can be introduced into the polynucleotide synthesis by replacing a portion of cDNA with one or more mutagenic oligodeoxynucleotides (polynucleotides) and transcribing the partially-mismatched template directly as has been previously described by Joyce et al., Nucleic Acid Research, 17:711-722 (1989). Amplification systems based on transcription have been described by Gingeras et al., in PCR Protocols, A Guide to Methods and Applications, pp 245-252, Academic Press, Inc., San Diego, CA (1990).

If the inducing agent is a DNA-dependent RNA polymerase and therefore incorporates ribonucleotide triphosphates, sufficient amounts of ATP, CTP, GTP and UTP are admixed to the primer extension reaction admixture and the resulting solution is treated as described above.

The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which can be used in the succeeding steps of the process.

The first and/or second PCR reactions discussed above can advantageously be used to incorporate into the receptor a preselected epitope useful in immunologically detecting and/or isolating a receptor. This is accomplished by utilizing a first and/or second polynucleotide synthesis primer or expression vector to incorporate a predetermined amino acid residue sequence into the amino acid residue sequence of the receptor.

After producing operatively linked V_H- and V_L-coding DNA homologs for a plurality of different V_H- and V_L-coding genes within the repertoires, the

dicistronic DNA molecules are typically further amplified. While the dicistronic DNA molecules can be amplified by classic techniques such as incorporation into an autonomously replicating vector, it is preferred to first amplify the molecules by subjecting them to a polymerase chain reaction (PCR) prior to inserting them into a vector. In fact, in preferred strategies, the first and second PCR reactions are performed in the same admixture that is subject to a multiplicity of PCR thermocycles where the outside primers are in molar excess. Preferably the number of PCR thermocycles is at least n+5, wherein n is the number of PCR thermocycles necessary to decrease by a factor of 10, and preferably exhaust, the number of inside primers by consumption in the formation of inside primer-primed products.

PCR is typically carried out by thermocycling i.e., repeatedly increasing and decreasing the temperature of a PCR reaction admixture within a temperature range whose lower limit is about 10°C to about 40°C and whose upper limit is about 90°C to about 100°C. The increasing and decreasing can be continuous, but is preferably phasic with time periods of relative temperature stability at each of temperatures favoring polynucleotide synthesis, denaturation and hybridization.

In preferred embodiments only one pair of first and second primers is used per amplification reaction. The amplification reaction products obtained from a plurality of different amplifications, each using a plurality of different primer pairs, are then combined.

However, the present invention also contemplates DNA homolog production via co-amplification (using two pairs of primers), and

multiplex amplification (using up to about 8, 9 or 10 primer pairs).

A diverse library of dicistronic DNA molecules having upstream and downstream cistrons can also be produced by combining, in a PCR buffer, double stranded V_H and V_L repertoires, V_H and V_L outside primers, and an inside primer having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal inside primer-template (primer-coding) portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion has a nucleotide base sequence homologous (identical) to a portion of the primer extension product of the other of the outside primers. That is, the linking primer has terminal sequences homologous to sequences in both repertoires. The cistronic bridge coding portion codes for, either directly or through complementarily, at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site located downstream from the stop codon, and a polypeptide leader having a translation initiation codon in the same reading frame as the downstream cistron, the initiation codon being located downstream from the ribosome binding site.

The dicistronic DNA molecules containing operatively linked V_H- and V_L-coding DNA homologs produced by PCR amplification are typically in double-stranded form and may have contiguous or adjacent to each of their termini a nucleotide sequence defining an endonuclease restriction site. Digestion of the dicistronic DNA molecules having restriction sites at or near their termini with one or more appropriate endonucleases results in the production of DNA

molecules having cohesive termini of predetermined specificity.

When individual PCR admixtures contain diverse gene repertoires the present invention produces many non-naturally occurring antibodies, i.e., combinations of V_H and V_L in a heterodimer. To take advantage of the mammalian immune system's capacity to select V_H and V_L combinations, the present invention also contemplates using fusion PCR to operatively link, and thereby recover, naturally occurring V_H and V_L combinations.

In preferred embodiments, a fusion PCR method of the present invention is performed on repertoires comprising a plurality of substantially isolated cells containing genes coding for a heterodimeric receptor. For example, a plurality of PCR admixtures is formed, each of which contains (i) a sample of substantially isolated B lymphocytes from a mammal producing antibody molecules against a preselected antigen, (ii) a PCR buffer, and (iv) either the previously described V_H and V_L PCR primer pairs or the set of outside V_H and V_L PCR primers in combination with the linking primer(s), also as previously described. The plurality of PCR admixtures is then subjected to a multiplicity of PCR thermocycles as described herein.

By "substantially isolated" is meant a sample containing less than about 100 target cells, such as B lymphocytes, T cells, and the like. In preferred embodiments, the plurality of PCR admixtures contain only about one cell. The cells are typically obtained from an individual mammal whose serum contains antibody molecules against the preselected antigen. The collected cells are typically seeded, usually at densities in the range of 0.5 to 100 cells per unit volume, into a plurality of individual PCR vessels,

such as microtiter plate wells and the like. Usually, the plurality of PCR admixtures is in the range of 800 to 1200, and preferably is about 1000, separate admixtures.

Typically, fewer cells are needed in each PCR admixture where the cells are obtained from individuals expressing a high serum antibody titer against the preselected antigen. For example, where B lymphocytes are obtained from an individual having a frequency of circulating B cells producing the antibody molecules of preselected specificity of 1/3000, each of about 800 to 1200 individual PCR admixtures need only contain about one B lymphocyte to result in isolation of the desired antibody. Where the circulating B cell frequency is in the range of 1/500,000, a density of about 100 cells per PCR admixture in each of about 800 to 1200 individual PCR admixtures will be needed before the process will result in isolation of the desired antibody.

In preferred embodiments, the PCR process is used not only to produce a library of dicistronic DNA molecules, but also to induce mutations within the library or to create diversity from a single parental clone and thereby provide a library having a greater heterogeneity. First, it should be noted that the PCR process itself is inherently mutagenic due to a variety of factors well known in the art. Second, in addition to the mutation inducing variations described in the above referenced U.S. Patent No. 4,683,195, other mutation inducing PCR variations can be employed. For example, the PCR reaction admixture, can be formed with different amounts of one or more of the nucleotides to be incorporated into the extension product. Under such conditions, the PCR reaction proceeds to produce nucleotide substitutions within

the extension product as a result of the scarcity of a particular base. Similarly, approximately equal molar amounts of the nucleotides can be incorporated into the initial PCR reaction admixture in an amount to efficiently perform X number of cycles, and then cycling the admixture through a number of cycles in excess of X, such as, for instance, 2X. Alternatively, mutations can be induced during the PCR reaction by incorporating into the reaction admixture nucleotide derivatives such as inosine, not normally found in the nucleic acids of the repertoire being amplified. During subsequent in vivo amplification, the nucleotide derivative will be replaced with a substitute nucleotide thereby inducing a point mutation.

4. Expressing the Dicistronic DNA Molecules

The dicistronic DNA molecules produced by the above-described method can be operatively linked to a vector for amplification and/or expression.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. One type of preferred vector is an episome, i.e., a nucleic acid molecule capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors".

The choice of vector to which a V_H- and V_L-coding DNA homolog is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., replication or

protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules. In preferred embodiments, the vector utilized includes a prokaryotic replicon i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra chromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a prokaryotic replicon also include a gene whose expression confers a selective advantage, such as drug resistance, to a bacterial host transformed therewith. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

Those vectors that include a prokaryotic replicon can also include a prokaryotic promoter capable of directing the expression (transcription and translation) of the V_H- and V_L-coding homologs in a bacterial host cell, such as E. coli transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur.

Promoters contain two highly conserved regions, one located about 10 bp (-10 region on Priberow box) and the other about 35 bp (-35 region) upstream from the point at which transcription starts. These two regions typically determine promoter strength. In addition, the number of nucleotides that separate the conserved sequences is important for efficient promoter function. For example, 16 to 19 nucleotides typically separate the -10 and -35 regions, and changes in that spacing can change the efficiency of a promoter. Promoter sequences

compatible with bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322, and pBR329 available from BioRad Laboratories, (Richmond, CA) and pPL and pKK223 available from Pharmacia, (Piscataway, NJ).

Promoters useful in this invention include Ptac ϕ 1.1A, ϕ 1.1B and ϕ10, which are recognized by T7 polymerase. See U.S. Patent No. 4,946,786. Useful regulatable promoters include the E. coli lac promoter described in U.S. Patent No. 4,946,786 and the promoters for the temperature sensitive genes in U.S. Patent No. 4,806,471. See also U.S. Patent No. 4,711,845.

A variety of methods have been developed to operatively link DNA to vectors via complementary cohesive termini. For instance, complementary cohesive termini can be engineered into the dicistronic DNA molecules during the primer extension reaction by use of an appropriately designed polynucleotide synthesis primer, as previously discussed. The dicistronic DNA molecule, and vector if necessary, is cleaved with a restriction endonuclease to produce termini complementary to those of the vector. The complementary cohesive termini of the vector and the dicistronic DNA molecule are then operatively linked (ligated) to produce a unitary double stranded DNA molecule.

The present method produces a diverse population of double stranded DNA expression vectors wherein each vector expresses, under the control of a single promoter, one V_H-coding DNA homolog and one V_L-coding DNA homolog, the diversity of the population being the result of different V_H- and V_L-coding DNA

homolog combinations that occurs during the PCR reaction where both outside and both inside primers are present in effective amounts. Preferably the vectors are linear double stranded DNA, such as a Lambda Zap derived vector as described herein.

In preferred embodiments, the vector defines a nucleotide sequence coding for a ribosome binding site and a leader, the sequence being located downstream from a promoter and upstream from a sequence coding for a polypeptide leader. In preferred embodiments, the vector contains a selectable marker such that the presence of a dicistronic DNA molecule of this invention inserted into the vector, can be selected. Typical selectable markers are well known to those skilled in the art. Examples of such markers are antibiotic resistance genes, genetically selectable markers, mutation suppressors such as amber suppressors and the like. The selectable markers are typically located upstream of the promoter.

The resulting construct is then introduced into an appropriate host to provide amplification and/or expression of the V_H- and V_L-coding DNA homologs. When coexpressed within the same organism, a functionally active heterodimeric receptor, such as an F_v, is produced. Cellular hosts into which a V_H- and V_L-coding DNA homolog-containing construct has been introduced are referred to herein as having been "transformed" or as "transformants".

The host cell can be either prokaryotic or eukaryotic. Bacterial cells are preferred prokaryotic host cells for library screening, and typically are a strain of E. coli such as, for example, the E. coli strain DH5 available from Bethesda Research Laboratories, Inc., Bethesda, MD. Preferred eukaryotic host cells include yeast and mammalian

cells, preferably vertebrate cells such as those from a mouse, rat, monkey or human cell line.

Transformation of appropriate cell hosts with a recombinant DNA molecule of the present invention is accomplished by methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al., Proc. Natl. Acad. Sci., USA, 69:2110 (1972); and Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY (1982). With regard to the transformation of vertebrate cells with retroviral vectors containing rDNAs, see for example, Sorge et al., Mol. Cell. Biol., 4:1730-1737 (1984); Graham et al., Virol., 52:456 (1973); and Wigler et al., Proc. Natl. Acad. Sci., USA, 76:1373-1376 (1979).

5. Screening For Expression of V_H and/or V_L Polypeptides

Successfully transformed cells, i.e., cells containing a dicistronic DNA molecule operatively linked to a vector, can be identified by any suitable well known technique for detecting the binding of a receptor to a ligand or the presence of a polynucleotide coding for the receptor, preferably its active site. Preferred screening assays are those where the binding of ligand by the receptor produces a detectable signal, either directly or indirectly. Such signals include, for example, the production of a complex, formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, cells from a population subjected to transformation with a subject recombinant DNA (rDNA) can be cloned to produce monoclonal colonies. Cells form those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA

using a method such as that described by Southern, J. Mol. Biol., 98:503 (1975) or Berent et al., Biotech., 3:208 (1985).

In addition to directly assaying for the presence of a dicistronic DNA molecule, successful transformation can be confirmed by well known immunological methods, especially when the V_H and/or V_L polypeptides produced contain a preselected epitope. For example, samples of cells suspected of being transformed are assayed for the presence of the preselected epitope using an antibody against the epitope.

Surface Expression

The present invention includes a method for expressing a polypeptide on the outer surface of E. coli. The surface expression of a polypeptide provides a particularly advantageous technique for screening diverse libraries for a polypeptide, such as a receptor, having a pre-selected activity. For example, E. coli expressing a diverse library of Fab fragments on their surface can be "panned" for transformants carrying antibody activity against a specific antigen.

E. coli surface expression is accomplished by fusing a portion of the lamB protein of E. coli to the polypeptide whose surface expression is desired. Any protein expressed on the cell surface of E. coli can provide the outer membrane spanning signal (surface expression signal) for use in the present invention. More specifically, it has been discovered that amino acid residues 57-181 of mature lamB can act as a signal for surface expression. Such fusion polypeptides are represented by the formula, shown in the direction of amino- to carboxy-terminus:

(F1) $NH_2 - B - Z - COOH,$

wherein B represents the amino acid residue sequence of a polypeptide, preferably heterologous to lamB, and Z represents a sequence of amino acid residues corresponding, and preferably identical, to the sequence of lamB from about residue position 57 to about residue position 181 as shown in Figure 3. The heterologous polypeptide can itself be a fusion protein, and typically contains a periplasmic secretion signal sequence (polypeptide leader), such as the pelB secretion signal, and the like. Thus, a preferred fusion polypeptide is represented by the formula,

(F2) $NH_2 - leader - J - Z - COOH$

wherein the leader is a sequence of amino acid residues that signal secretion to the periplasm, J is a sequence of amino acid residues of from 6 to 350 residues in length, and Z is as described before in formula (F1). Preferably J is from about 50 to about 150 amino acid residues. More preferably, J is a V_H or V_L as described herein.

Recombinant DNA Molecules

In view of the foregoing, the present invention contemplates. In living organisms, the amino acid residue sequence of a protein or polypeptide is directly related via the genetic code to the deoxyribonucleic acid (DNA) sequence of the structural gene that codes for the protein. Thus, a structural gene for a fusion protein of this invention can be defined in terms of the amino acid residue sequence, i.e., protein or polypeptide, for which it

codes. In addition an important and well known feature of the genetic code is its redundancy. That is, for most of the amino acids used to make proteins, more than one coding nucleotide triplet (codon) can code for or designate a particular amino acid residue. Therefore, a number of different nucleotide sequences may code for a particular amino acid residue sequence. Such nucleotide sequences are considered functionally equivalent since they can result in the production of the same amino acid residue sequence in all organisms. Occasionally, a methylated variant of a purine or pyrimidine may be incorporated into a given nucleotide sequence. However, such methylations do not affect the coding relationship in any way.

Recombinant DNA molecules containing a nucleic acid sequence coding for a fusion polypeptide according to formulas (F1) or (F2) are contemplated by this invention. Thus, the present invention provides for linking a nucleotide sequence coding for any polypeptide immunogen against which antibody production is desired to the outer membrane spanning signal (lamB) polypeptide and/or the secretion signal (pel B) polypeptide as described herein. In preferred embodiments the polypeptide immunogen is a pathogen related immunogen and the conjugate has the capacity to induce the production of antibodies that immunoreact with the pathogen when injected in an effective amount into an animal. Exemplary immunogens of particular importance are derived from bacteria such as B. pertussis, S. typhosa, S. paratyphoid A and B, C. diptheriae, C. tetani, C. botulinum, C. perfringens, B. anthracis, P. pestis, P. multocida, V. cholerae, N. meningitides, N. gonorrhea, H. influenzae, T. palladium, and the like; immunogens derived from viruses such as polio virus, adenovirus,

parainfluenza virus, measles, mumps, respiratory syncytical virus, influenza virus, equine encephalomyeitis virus, hog cholera virus, Newcastle virus, fowl pox virus, rabies virus, pseudorabies virus, feline and canine distemper viruses and the like; rickettsiae immunogen such as epidemic and endemic typhus, and the spotted fever groups, and the like. Immunogens are well known to the prior art in numerous references such as U.S. Patent No. 3,149,036, No. 3,983,228, and No. 4,069,313; in Essential Immunology, 3rd Ed., by Roit, published by Blackwell Scientific Publications; in Fundamentals of Clinical Immunology, by Alexander and Good, published by W.B. Saunders; and in Immunology, by Bellanti, published by W.B. Saunders.

Methods for determining the presence of antibodies to an immunogen in a body sample from an immunized animal are well known in the art.

In preferred embodiments the polypeptide immunogen is a pathogen related immunogen that immunoreacts with, i.e., is immunologically bound by, antibodies induced by the pathogen. More preferably, the pathogen related immunogen is capable of inducing an antibody response that provides protection against infection by the pathogen. Methods for determining the presence of both cross-reactive and protective antibodies are well known in the art.

Expression Vectors

The present invention also contemplates various expression vectors useful in performing, inter alia, the methods of the present invention. Each of the expression vectors is a novel derivative of Lambda Zap.

1. Lambda Zap II

Lambda Zap II is prepared by replacing the Lambda S gene of the vector Lambda Zap with the Lambda S gene from the Lambda gt10 vector, as described in Example 7.

2. Lambda ImmunoZAP H

Lambda ImmunoZAP H is prepared by inserting the synthetic DNA sequences illustrated in Figure 6A into the above-described Lambda Zap II vector. The inserted nucleotide sequence advantageously provides a ribosome binding site (Shine-Dalgarno sequence) to permit proper initiation of mRNA translation into protein, and a leader sequence to efficiently direct the translated protein to the periplasm. The preparation of Lambda ImmunoZAP H is described in more detail in Example 8, and its features illustrated in Figures 6A and 7.

3. Modified Lambda ImmunoZAP H

Modified Lambda ImmunoZAP H is prepared by inserting the modified synthetic DNA sequences illustrated in Figure 8A into the above-described Lambda ZAP II vector. The preparation of modified Lambda ImmunoZAP H and the details of the modifications are described in Example 8B. Its features are illustrated in Figure 8A and 8B.

4. Lambda ImmunoZAP L

Lambda ImmunoZAP L is prepared as described in Example 9 by inserting into Lambda Zap II the synthetic DNA sequence illustrated in Figure 6B. Important features of Lambda ImmunoZAP L are illustrated in Figure 9.

Transformants and Vaccines

A host transformed with a recombinant DNA molecule of this invention is also contemplated by this invention. The transformants are useful, not

only in isolating a heterodimeric receptor according to the methods described herein, but also as vaccine strains where the fusion polypeptide immunologically cross-reacts with pathogen-neutralizing antibodies. Methods of formulating and using vaccine strains are described in U.S. Patents No. 4,764,370 and No. 4,337,314.

For heterodimeric molecules that assemble in the cell or in the periplasm, operatively linking the lamB outer membrane spanning signal sequence to the carboxy-terminus of one of the polypeptide chains of the heterodimer, e.g., the heavy chain of a Fab, results in surface expression of the assembled heterodimer.

One of the advantages of the present invention is that a vaccine containing a transformant of this invention can be easily prepared, lyophilized in the presence of appropriate inert, non-toxic carrier(s) (infra) in vials and stored at room temperature without loss of potency. No refrigeration or special storage equipment is required.

The composition of vaccine preparations must be known and consistent. This achieved by using specified amounts of quality-controlled chemical and biological ingredients in their preparation. Methods for the quality control of chemical components are well established in the art and will not be discussed here. Chemical purity in the vaccine preparations is defined as freedom from toxic waste or cellular breakdown products and interfering or spurious immunogenic material. This is assured by working with pure cultures (the vaccine strain free of other cells or virus) and harvesting the cells while the culture is in the logarithmic phase of growth (before the synthesis of autolytic enzymes). The collection and

washing of cells from the medium by physical methods (centrifugation) should leave low molecular weight impurities in the supernatant.

The vaccines of the present invention can be administered to any warm-or cold-blooded animals susceptible to infection with pathogenic microorganisms. Human and non-human animals may benefit as hosts.

Administration can be parenteral, but preferably oral or intranasal, depending upon the natural route of infection. In farm animals, for example, the vaccine may be administered orally, by incorporation of the vaccine in feed or feed water. The dosage administered may be dependent upon the age, health and weight of the recipient, kind of concurrent treatment if any, and nature of the organism. Generally, a dosage of active ingredient will be from about 10^1 to 10^{10} cells per application per host. The preferred dose for intranasal administration would generally be about 10^6 organisms, suspended in 0.05 to 0.1 ml of an immunologically inert carrier. Peroral administration of a vaccine strain of, for example, Salmonella typhi developed according to the method described in this invention would probably require 10^6 to 10^8 organisms suspended in 1-2 mls of, for example, skim milk. The vaccines can be employed in dosage forms such as capsules, liquid solutions, suspensions, or elixirs, for oral administration, or sterile liquid for formulations such as solutions or suspensions for parenteral, intranasal or topical (e.g. wounds or burns) use. An inert, immunologically acceptable carrier is preferably used, such as saline, phosphate buffered saline or skim milk.

Compositions and Kits

Many of the reagents described herein (e.g., nucleic acids such as primers, vectors, and the like) have a number of forms, particularly variably protonated forms, and in equilibrium with each other. As the skilled practitioner will understand, representation herein of one form of a compound or reagent is intended to include all forms thereof that are in equilibrium with each other.

The reagents described herein can be packaged in kit form. As used herein, the term "package" refers to a solid matrix or material customarily utilized in a system and capable of holding within fixed limits one or more of the reagent components for use in a method of the present invention. Such materials include glass and plastic (e.g., polyethylene, polypropylene and polycarbonate) bottles, vials, paper, plastic and plastic-foil laminated envelopes and the like. Thus, for example, a package can be a glass vial used to contain the appropriate quantities of polynucleotide primer(s), vectors, restriction enzyme(s), DNA polymerase, DNA ligase, or a combination thereof. An aliquot of each component sufficient to perform at least one PCR thermocycle will be provided in each container.

Kits useful for producing a template-complement or for amplification of a specific nucleic acid sequence using a primer extension reaction methodology also typically include, in separate containers within the kit, dNTPs where N is adenine, thymine, guanine and cytosine, and other like agents for performing primer extension reactions.

The reagent species of any system described herein can be provided in solution, as a liquid dispersion or as a substantially dry powder, e.g., the plasmids may be provided in lyophilized form.

In one embodiment, the kit is an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L gene primer, and a linking primer defining a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer template portion. The 3'-terminal priming portion has a nucleotide base sequence complementary to a portion of the primer extension product of one of the outside primers. The 5'-terminal primer-template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of the outside primers. The cistronic bridge coding portion is as previously described.

Another contemplated kit comprises an enclosure containing, in separate containers, an outside first polypeptide, preferably a V_H, gene primer, an outside second polypeptide, preferably a V_L, gene primer, an inside first polypeptide, preferably a V_H, gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion. The 3'-terminal priming portion comprises a nucleotide sequence homologous to a conserved portion of a V_H gene. The kit also contains an inside second polypeptide, preferably a V_L, gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of the first polypeptide gene primer, the 3'-terminal priming portion of which comprises a nucleotide sequence homologous to a conserved portion of a second polypeptide gene. The first polypeptide inside and second polypeptide inside primers, when hybridized, form a duplex that codes for a double-stranded DNA molecule containing the before described cistronic bridge for linking the upstream and

downstream cistrons.

Examples

The following examples are intended to illustrate, but not limit, the present invention.

1. Oligonucleotide Primer Design for Producing Dicistronic DNA

A method based on PCR amplification that fuses heavy and light chain sequences has been used to construct a complete antigen binding domain of a Fab protein fragment composed of a heavy and a light chain. Schematic diagrams of an immunoglobulin molecule composed of heavy and light chains containing constant and variable regions is shown in Figure 1. Human heavy chain IgG and human kappa light chain are diagrammatically sketched in Figures 2A and 2B, respectively. To accomplish this procedure, immunoglobulin heavy and light chain primers were designed to produce a region of homology between two polymerase chain reaction (PCR) products. The complementary regions have been shown to hybridize predominantly under conditions where one set of primers ("inside primer pair") is used in a limiting amount relative to the other set of primers ("outside primer pair"). After the 3' ends of the PCR products have hybridized, the DNA polymerase has been shown to extend the ends creating a fusion sequence carrying the unique sequences of both PCR fragments separated by one copy of region X cistronic bridge. A two-step cloning procedure is thus avoided. When the recombinant sequence is then inserted into an expression vector such as ImmunoZAP, a fusion product capable of simultaneously expressing the heavy and light chains can be produced.

The strategy used for producing immunoglobulin heavy and light chain PCR dicistronic DNA is shown schematically in Figure 4. Regions of the immunoglobulin heavy chain coding strand are designated V_H, C_H1, C_H2, and C_H3 corresponding to functional regions in the protein. The corresponding regions of the non-coding strand are designated by a prime ('). Regions V_L and C_L are similarly labelled for the kappa light chain. This procedure can also be performed using lambda light chain specific regions. A region, X, unrelated to the natural immunoglobulin sequences, is introduced into the fusion product by attaching X to the 5' ends of both of the C_H1' and V_L inside primers.

Overlapping oligonucleotide primers used in the fusion-PCR reactions to produce dicistronic DNA were designed to encode the following: amino acids of 225 to 230 of the IgG heavy chain hinge region which are common to all human IgG isotypes; an Spe I restriction site; two stop codons; a ribosome binding site; a periplasmic (pelB) leader sequence (Better, et al., Science, 240:1041-1043 (1988); Lei, et al., J. Bacteriol., 169: 4379-4383 (1988)); a Sac I restriction site which encodes amino acids 1 and 2 of the mature kappa light chain; and amino acids 3 to 8 of the mature kappa light chain. The X region was designed to contain a ribosome binding site and a pelB leader to ensure expression of the light chain. Nucleotide sequences for all human and mouse PCR primers, both inside and outside, are listed in Table 7. Primers followed by a prime (') represent non-coding strand sequences.

Table 7

Human and Mouse PCR Primers

	Seq. Id. No.		
		Human	
	(56)	V_H	5'-GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG-3'
	(57)	C_H1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGGT ATTTCATTATGACTGTCTCCTTGCTATTAACTAG TACAAGATTTGGGCTC-3'
	(58)	V_L	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGCT GCCCAACCTGCCATGGCTGAGCTCGTGATGACCC CAGTCTCC-3'
	(59)	C_L'	5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA GCTCTTTGTGACGGGCGAACTC-3'
		Mouse	
	(60)	V_H	5'-AGGTCCAGCTGCTCGAGTCTGG-3'
	(61)	C_H1'	5'-AATAACAATCCAGCGGCTGCCGTAGGCAATAGG TATTTCATTATGACTGTCTCCTTGCTATTAACT AGTATACAATCCCTGGGCACAAT-3'
	(62)	V_L	5'-GCCTACGGCAGCCGCTGGATTGTTATTAATCGC TGCCCAACCTGCCATGGCTGAGCTCGTGATGAC CCAGTCTCC-3'
	(63)	C_L'	5'-TCCTTCTAGATTACTAACACTCTCCCCTGTTGAA-3'

The overlapping regions of the human C_H1' inside and V_L inside primers are illustrated in Figure 5. The heavy chain downstream C_H1' inside primer sequence is written 3' to 5' and the light chain upstream V_L inside primer sequence is written 5' to 3'. The complementary PCR product strands, and not the primer strands, cross-prime to create the dicistronic molecule. Bold nucleotides represent regions where the C_H1' inside primer hybridizes to the 3' end of C_H1 on human IgG heavy chain mRNA or where

the V_L inside primer hybridizes to the 5' end of V_L framework on human kappa light chain cDNA. The amino acid and nucleotides in italics represent changes in sequence from the original pelB leader sequence.

At amino acid 15 of the pelB leader sequence, the codon was changed from CTC to ATC resulting in a conservative amino acid change from a leucine to an isoleucine as shown in Figure 5 and Table 7. Hydrophobic amino acids in the core region of periplasmic leader sequences have been shown to be essential for correct processing of the leader sequence and transport of the mature protein to the periplasm. Oliver, in Neidhardt, R.C. (ed.), Escherichia coli and Salmonella Typhimurium., Am. Soc. Microbiol., 1:56-69 (1987). The nucleotide changes were made to allow for the artifactual insertion of one or two dATPs at the 3' end of the overlapping dicistronic molecules. Thermus aquaticus (Taq) DNA polymerase may add a dATP to the 3' end of the PCR product because of terminal transferase activity. Jiang, et al. Oncogene, 4: 923-928 (1989). The additional dATP would then cause a mismatch between the overlapping PCR products at the 3' terminus and inhibit elongation by Taq DNA polymerase. Sommer, et al. Nucl. Acids Res., 17: 6749 (1989). Therefore, the change to two dTTPs in this position of the oligonucleotide primers would allow proper base pairing if up to two dATPs were added to the 3' terminus of the heavy chain PCR product. The kappa light chain PCR product was designed to terminate at a position where two dTTPs occur 5' of the end of the product and did not require alterations of the nucleotide sequence. Nucleotides were changed in the kappa light chain primer encoding the pelB leader sequence without introducing amino acid changes in

order to decrease the number of mismatches between the primer and the leader sequence of the kappa light chain mRNA as shown in Figure 5 and Table 7.

All primers were synthesized on an Applied Biosystems DNA synthesizer, model 381A, following the manufacturer's instructions.

2. Preparation of a V_H-and V_L-Coding Repertoire

a. Preparation of a V_H-and V_L-Coding Repertoire from a Human cDNA Combinatorial Library

Cloned DNA, previously isolated from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT) was used as a template for preparing a V_H-and V_L-coding repertoire. Mullinax, et al., supra. Briefly, the combinatorial library was prepared by the following approach. Volunteer donors, who had been previously immunized against tetanus but had not received booster injections within the last year, received injections on 2 consecutive days of 0.5 milliliters (ml) of alum-absorbed TT (40 microgram/ml (ug)/ml) (Connaught Laboratories, Swiftwater, Pennsylvania).

One hundred ml of blood was drawn from the volunteers 6 days post injection and anticoagulated with a mixture of 0.14 M citric acid, 0.2 M trisodium citrate, and 0.22 M dextrose. The peripheral blood lymphocytes (PBLs) were recovered and isolated from the whole blood by layering the whole blood on Histopaque-1077 (Sigma, St. Louis, Missouri) and centrifuging at 400 x g for 30 minutes at 25 degrees Celsius (25°C). Isolated PBLs were washed twice with phosphate buffered saline (PBS) (150 mM sodium chloride and 150 mM sodium phosphate, pH 7.2 at 25°C).

Total RNA was then purified from the PBLs (10^6 B cells per ml blood per 100 ml of blood) for an enriched source of B-cell mRNA coding for anti-TT IgG using an RNA isolation kit according to manufacturer's instructions (Stratagene, La Jolla, California) and also described by Chomczynski et al., Anal. Biochem., 162:156-159 (1987). Briefly, the isolated PBLs were homogenized in 10 ml of a denaturing solution containing 4.0 M guanine isothiocyanate, 0.25 M sodium citrate at pH 7.0, and 0.1 M beta-mercaptoethanol. One ml of sodium acetate at a concentration of 2 M at pH 4.0 was admixed with the homogenized cells. Ten ml of phenol that had been previously saturated with H_2O was also admixed to the denaturing solution containing the homogenized cells. Two ml of a chloroform:isoamyl alcohol (24:1 v/v) mixture was added to this homogenate. The homogenate was mixed vigorously for ten seconds and maintained on ice for 15 minutes. The homogenate was then transferred to a thick-walled 50 ml polypropylene centrifuged tube (Fisher Scientific Company, Pittsburgh, Pennsylvania). The solution was centrifuged at 10,000 x g for 20 minutes at 4°C. The upper RNA-containing aqueous layer was transferred to a fresh 50 ml polypropylene centrifuge tube and mixed with an equal volume of isopropyl alcohol. This solution was maintained at -20°C for at least one hour to precipitate the RNA. The solution containing the precipitated RNA was centrifuged at 10,000 x g for twenty minutes at 4°C. The pelleted total cellular RNA was collected and dissolved in 3 ml of the denaturing solution described above. Three ml of isopropyl alcohol was added to the re-suspended total cellular RNA and inverted to mix. This solution was maintained at -20°C for at least 1 hour to precipitate the RNA. The solution containing the precipitated RNA

was centrifuged at 10,000 x g for ten minutes at 4°C. The pelleted RNA was washed once with a solution containing 75% ethanol. The pelleted RNA was dried under vacuum for 15 minutes and then re-suspended in diethyl pyrocarbonate (DEPC) treated (DEPC-H_2O) H_2O.

Messenger RNA (mRNA) was prepared from the total cellular RNA using methods described in Molecular Cloning A Laboratory Manual, Maniatis et al., eds., Cold Spring Harbor, NY, (1982). Briefly, 500 mg of the total RNA isolated from a PBLs prepared as described above was re-suspended in one ml of 1X sample buffer (1 mM Tris-HCl, (Tris [hydroxylmethyl-aminomethane]) pH 7.5; 0.1 mM EDTA (disodium ethylene diamine tetra-acetic acid), 0.5 M NaCl) and maintained at 65°C for five minutes and then on ice for five more minutes. The mixture was then applied to an oligo-dT (Stratagene) column that was previously prepared by washing the oligo-dT with a solution containing 10 mM Tris-HCl, pH 7.5; 1 mM EDTA, 0.5 M NaCl. The eluate was collected in a sterile polypropylene tube and reapplied to the same column after heating the eluate for five minutes at 65°C. The oligo dT column was then washed with 0.4 ml of high salt loading buffer consisting of 10 mM Tris-HCl at pH 7.5, 500 mM sodium chloride, and 1 mM EDTA. The oligo dT column was then washed with 2 ml of 1 X low salt buffer consisting of 10 mM Tris-HCl at pH 7.5, 100 mM sodium chloride, and 1 mM EDTA. The messenger RNA was eluted from the oligo dT column with 0.6 ml of buffer consisting of 10 mM Tris-HCl at pH 7.5, and 1 mM EDTA. The messenger RNA was purified by extracting this solution with phenol/chloroform followed by a single extraction with 100% chloroform. The messenger RNA was concentrated by ethanol precipitation and re-suspended in DEPC H_2O.

The messenger RNA isolated by the above process contains a plurality of different V_H and V_L coding polynucleotides, i.e., greater than about 10^4 different V_H- and V_L-coding genes.

Isolated RNA was converted to cDNA by a primer extension reaction with a first-strand synthesis kit according to manufacturer's instructions (Stratagene) by using an oligo (dT) primer for the light chain and a specific primer, C_H1', for the heavy chain. Mullinax et al., supra. In a typical 50 μl transcription reaction, 5 ug of PBL mRNA in water was first hybridized (annealed) with 200 ng (50.0 pmol) of an oligo (dT) primer for the light chain. In a separate reaction, 5 ug of PBL mRNA in water was first hybridized (annealed with 200 ng (20 pmol) of the heavy chain primer, C_H1', at 65°C for five minutes. Subsequently, the mixture was adjusted to 0.5 mM each of dATP, dCTP, dGTP and dTTP, 50 mM Tris-HCl at pH 8.3, 3 mM $MgCl_2$, 75 mM KCl, 10 mM DTT, 20 units of RNase block II (Stratagene), and 20 units of Moloney-Murine Leukemia virus reverse transcriptase (Stratagene Cloning Systems), was added and the solution was maintained for 1 hour at 37°C. PCR amplification of the heavy and light chain sequences was done separately using 0.25-0.5 ug of first-strand synthesis product as template with sets of primer pairs using Taq DNA polymerase as described in Example 3.

The PCR amplified light chain DNA fragments were then digested with Sac I and Xba I and ligated into a modified Lambda Zap II vector as prepared in Example 9 to form a light chain ImmunoZap Library (ImmunoZAP L; Stratacyte, La Jolla, California). The PCR amplified heavy chain DNA was digested with Spe I and Xho I and ligated into a different modified Lambda

Zap II vector as prepared in Example 7 to form a heavy chain ImmunoZap Library (ImmunoZAP H; Stratacyte). The resulting libraries were amplified and the resulting DNA was packaged into bacteriophage with *in vitro* packaging extract, Gigapack II gold (Stratagene) and used to infect *E. coli* strain XL1-Blue (Stratagene).

To construct a library for coexpression, the right arm of the heavy chain library phage DNA was digested with Hind III, preserving the left arm of ImmunoZAP H with heavy chain inserts. The left arm of the light chain library phage DNA was digested with Mlu I resulting in a right arm of ImmunoZAP with kappa light chain inserts. Both products were then digested with Eco RI and ligated to create a combinatorial library that encoded human Fab fragments including those specific for TT. Mullinax, et al., *supra*.

Reactive plaques were first identified by binding to tetanus toxoid as described in Example 11. Bacteriophage from purified reactive plaques were then converted to the plasmid format by *in vivo* excision with R408 helper phage (Stratagene) following methods described in Example 11 and familiar to one skilled in the art. Short, et al., *Nucl. Acids. Res.*, 16:7583-7600 (1988). The resulting purified plasmid DNA encoding heavy and light chain was then used in PCR reactions as described below in Example 3.

b. Preparation of a V_H- and V_L-Coding Repertoire from mRNA from Tissues and Cells

1) Human

Purified populations of PBLs, other lymphocytes, and hybridomas which express immunoglobulins including IgG, IgM, IgE, IgD, and IgA are used as sources for isolating mRNA encoding

immunoglobulins. PBL's and other immunoglobulin expressing lymphocytes are isolated from either spleen, lymphoid tissue or plasma. Following purification of the cells, total RNA is then purified from these cells using a RNA isolation kit (Stratagene) as described in Example 2a. The purified RNA is then converted to cDNA with a first-strand synthesis kit as described in Example 2a. The resultant cDNA is then used as a template in PCR amplication reactions as described below in Example 3 for the production of dicistronic molecules expressing heavy and light chains.

2) Mouse

Populations of cells described above can be isolated from other mammalian sources such as mouse or rabbit. Both mRNA and rearranged DNA can be isolated as described above and used as templates in PCR amplification reactions. cDNA synthesized from mRNA isolated from a mouse anti-human fibronectin hybridoma (ATCC, CRL-1606) was used as a preferred template for the production of dicistronic molecules expressing heavy and light chain.

c. Preparation of a V_H-Coding Repertoire From Rearranged DNA

Rearranged DNA isolated from PBLs, other lymphocytes, and hybridomas which express immunoglobulins can be used to prepare a V_H-coding repertoire. The amplification procedure for preparing a V_H-coding repertoire using rearranged DNA is performed as described in Example 3.

3. Preparation of DNA Homologs

a. V_H-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 2 from a combinatorial library that encodes human Fab fragments

which bind tetanus toxoid (TT), was used as a template for preparing a V_H-coding double stranded DNA homolog. Human heavy chain, containing both the V_H and C_H1 coding region and designated as Fd, was amplified in a PCR reaction. The amplification was performed in a 100 ul reaction containing 5 nanograms (ng) of the cloned DNA in PCR buffer consisting of the following: 10 mM Tris-HCl, pH 8.3; 50 mM KCl, 1.5 mM $MgCl_2$; 0.001% (w/v) gelatin; 200 mM of each dNTP; 200 nanomolar (nM) of each primer; and 2.5 units of Taq DNA polymerase. The human V_H outside primer and C_H1' inside primer were used as a PCR primer pair for amplification of the heavy chain (Table 7 and Figure 4). The reaction mixture was overlaid with mineral oil and subjected to 40 cycles of amplification. Each amplification cycle (thermocycle) involved denaturation at 94°C for 1.5 minutes, annealing at 54°C for 2.5 minutes and polynucleotide synthesis by primer extension (elongation) at 72°C for 3.0 minutes followed by a return to the denaturation temperature. The resultant amplified V_H-coding DNA homolog containing samples were then gel purified, extracted twice with phenol/chloroform, once with chloroform followed by ethanol precipitation and were stored at -70°C in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA.

To verify the amplification of the heavy chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the heavy chain was approximately 730 base pairs as shown in Figure 10. The V_H-coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with V_L-coding counterparts prepared below for the production of dicistronic DNA molecules having V_H and V_L cistronic portions as illustrated in Example 4.

b. V_L-Coding Double Stranded DNA Homologs

Cloned DNA, prepared in Example 2 from a combinatorial library that encodes human Fab fragments which bind tetanus toxoid (TT), was used as a template for preparing a V_L-coding double stranded DNA homolog. Human light chain, containing the entire coding region of kappa light chain (V_L and C_L), was amplified using the same PCR conditions described for human heavy chain with the exception that a human V_L inside primer and C_L' outside primer were used as the PCR primer pair (Table 7 and Figure 4). The resultant V_L-coding double stranded DNA homolog was gel purified and stored as described above.

To verify the amplification of the light chain, the PCR purified products were electrophoresed in an agarose gel. The expected size of the light chain was approximately 690 base pairs as shown in Figure 10. The V_L-coding double stranded DNA homologs were then used in subsequent PCR amplification reactions with V_H-coding counterparts prepared above for the production of dicistronic DNA molecules as illustrated in Example 4.

4. Preparation of Internally-Primed Duplexes of V_H- and V_L-Coding DNA Homologs

a. Hybridization of V_H- with V_L-Coding DNA Homologs

The V_H- and V_L-coding double stranded DNA homologs prepare in Example 3a and 3b, respectively, were admixed together and denatured at 95°C for 5 minutes to separate the strands of each homolog. The denatured V_H-and V_L-coding DNA strands in the admixture were then annealed at 54°C for 5 minutes to form a V_H- and V_L-coding duplex DNA molecule hybridized at the 3' ends at region X of each original

homolog. One strand of the X region (cistronic) bridge encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon, and a polypeptide leader (pelB) having a translation initiation codon in the same reading frame as the downstream cistron located downstream from the ribosome binding site.

b. Primer Extension to Produce Dicistronic DNA Molecules

The hybridized recombinant V_H- and V_L-coding DNA molecule (internally primed duplex) was subjected to primer extension and then amplified with only the V_H and C_L' primers following the PCR reaction procedure described in Example 3a. This second PCR reaction is schematically represented in Figure 4. The PCR reaction products were gel electrophoresed to verify the presence of the resultant V_H-and V_L-coding dicistronic DNA molecules. The expected size of the dicistronic molecule was about 1390 base pairs and is shown in Figure 10. The resultant V_H- and V_L-coding dicistronic DNA molecules were then ligated into the modified ImmunoZAP H vector (Figure 8) for the construction of expression vectors as described in Example 10.

5. Preparation of Mouse Hybridoma V_H- and V_L-Coding Double Stranded DNA Homologs and Production of Dicistronic DNA Molecules in a Single Amplification Reaction

Mouse hybridoma heavy and light chain cDNA prepared in Example 2b was amplified in a single PCR reaction using the reaction conditions given above with an excess of the outside primers (200 nM concentration of both the mouse V_H primer and C_L'

primer) and a limiting amount of the inside primers (20 nM concentration of both the mouse C_H1' and V_L primer) (Table 7). The resultant mouse heavy and light chain dicistronic molecules were then inserted into a modified ImmunoZAP H for construction of an expression vector as described in Example 10.

6. Preparation of Internally-Primed Duplexes Using a Single Internal Primer that Overlaps Both the V_H and V_L Repertoires

Another approach to producing a library of dicistronic DNA molecules is to use a single internal primer instead of using two separately internal primers. The process of creating a dicistronic molecule comprising an upstream V_H cistron and a downstream V_L cistron is to combine in a PCR buffer the following: a repertoire of V_H genes consisting of at least 10^5 different genes; a repertoire of V_L genes consisting of at least 10^4 different genes; an outside V_H primer; an outside V_L primer; and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion. The PCR reaction is performed as described in Example 2a.

The 3'-terminal priming portion of a polynucleotide strand (linker) has a nucleotide base sequence homologous to a portion of the primer extension product of one of the outside primers. The 5'-terminal priming portion encodes a nucleotide base sequence homologous to a portion of the primer extension product of the other outside primer. The cistronic bridge coding portion encodes at least one stop codon in the same reading frame as the upstream cistron, a ribosome binding site downstream from the stop codon and a polypeptide leader (pelB) having a

translation initiation codon in the same reading frame as the downstream cistron where the initiation codon is located downstream from the ribosome binding site. Polynucleotide strand (linker) primers useful in this invention are listed in Table 8.

Table 8

Polynucleotide Strand (Linker) Primers

Seq. Id. No.			
(64)[1]	1'	5'	GGAGAGTGGGTCATCACGAGCTCAGCCATGGCAGGTTGG GCAGCGATTAATAACAATCCAGCGGCTGCCGTAGGCAAT AGGTATTTCATTATGACTGTCTCCTTGCTATTAACTAGT ACAAGATTTGGGCTC 3'
(65)[2]	2'	5'	GAGCCCAAATCTTGTACTAGTTAATAGCAAGGAGACAGT CATAATGAAATACCTATTGCCTACGGCAGCCGCTGGATT GTTATTAATCGCTGCCCAACCTGCCATGGCTGAGCTCGT GATGACCCACTCTCC 3'

[1] Primes mRNA (sense strand) of heavy chain C_H1 region; antisense strand of light chain V_L with dicistronic bridge in between heavy and light chains will be in the same relative orientation as given in the example.

[2] Primes antisense strand of heavy chain C_H1 regions; and sense strand of light chain V_L region with dicistronic in between heavy and light chains will be in the same relative orientation as given in the example.

The resultant single step internally primed dicistronic DNA molecule can then be ligated into modified ImmunoZAP H for construction of an expression vector as described in Example 10.

7. Preparation of Lambda Zap II Expression Vector

The vector Lambda Zap™ II (Stratagene) is a derivative of the original Lambda Zap (ATCC # 40,298) that maintains all of the characteristics of the original Lambda Zap including 6 unique cloning sites, fusion protein expression, and the ability to rapidly excise the insert in the form of a phagemid (Bluescript SK-), but lacks the SAM 100 mutation, allowing growth on many Non-Sup F strains, including XL1-Blue. The Lambda Zap II was constructed as described in Short et al., Nucleic Acids Res., 16:7583-7600, 1988, by replacing the Lambda S gene contained in a 4254 base pair (bp) DNA fragment produced by digesting Lambda Zap with the restriction enzyme NcoI. This 4254 bp DNA fragment was replaced with the 4254 bp DNA fragment containing the Lambda S gene isolated from Lambda gt10 (ATCC # 40,179) after digesting the vector with the restriction enzyme NcoI. The 4254 bp DNA fragment isolated from lambda gt10 was ligated into the original Lambda Zap vector using T4 DNA ligase and standard protocols for such procedures described in Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, NY, 1987.

8. Preparation of V_H-Expression Vectors, ImmunoZAP H and Modified ImmunoZAP H, Construction

a. ImmunoZAP H

The main criterion used in choosing a vector system was the necessity of generating the largest number of Fab fragments which could be screened directly. Bacteriophage lambda was selected as the expression vector for three reasons. First, in vitro packaging of phage DNA is the most efficient method of reintroducing DNA into host cells. Second,

it is possible to detect protein expression at the level of single phage plaques. Finally, the screening of phage libraries typically involve less difficulty with nonspecific binding. The alternative, plasmid cloning vectors, are only advantageous in the analysis of clones after they have been identified. This advantage is not lost in the present system because of the use of lambda Zap, thereby permitting a plasmid containing the heavy chain, light chain, or Fab expressing inserts to be excised.

To express the plurality of V_H-coding DNA homologs in an E. coli host cell, a vector was constructed that placed the V_H-coding DNA homologs in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, 1975, provided a leader sequence directing the expressed protein to the periplasmic space, provided a polynucleotide sequence that coded for a known epitope (epitope tag) and also provided a polynucleotide that coded for a spacer protein between the V_H-coding DNA homolog and the polynucleotide coding for the epitope tag. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6A. The individual single-stranded polynucleotides (N_1-N_{12}) are shown in Table 9 below.

Table 9

Seq. Id. No.		
(66)	N1)	5' GGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
(67)	N2)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(68)	N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'

(69) N4) 5' AGGTGAAACTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
(70) N5) 5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(71) N6) 5' CAGTTTCACCTGGGCCATGGCTGGTTGGG 3'
(72) N7) 5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(73) N8) 5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGC 3'
(74) N9-4) 5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(75) N11) 5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(76) N12) 5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(77) N10-5) 5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

Polynucleotides 2, 3, 9-4', 11, 10-5', 6, 7 and 8 were kinased by adding 1 μl of each polynucleotide (0.1 ug/μl) and 20 units of T_4 polynucleotide kinase to a solution containing 70 mM Tris-HCl at pH 7.6, 10 mM $MgCl_2$, 5 mM DTT, 10 mM beta mercaptoethanol, 500 ug/ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides, 20 ng, of polynucleotides N1 and polynucleotides N12, were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all 10 polynucleotides annealed to form the double stranded synthetic DNA insert shown in Figure 6A. The individual polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert by adding 40 μl of the above reaction to a solution containing 50 mM Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining

the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 µl of the above reaction, 4 µl of a solution containing 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a lambda Zap II vector prepared in Example 7 that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract (Stratagene). The packaged ligation mixture was plated on XL1-blue cells (Stratagene). Individual Lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol provided by the manufacturer (Stratagene). This in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a plasmid vector to allow easy manipulation and sequencing. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described in by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacture's instructions in the AMV Reverse Transcriptase ^{35}S-ATP sequencing kit (Stratagene). The sequence of the resulting V_H expression vector is shown in Figure 6A and Figure 7.

b. Modified ImmunoZAP H

To create a fusion-PCR library from hybridoma RNA for expressing the plurality of V_H-coding DNA homologs in an E. coli host cell, a vector based on the ImmunoZAP H vector described above was constructed. The procedure for constructing the

vector was performed as described above with the following modifications: elimination of the Sac I site between the T_3 polymerase and Not I sites and changing the nucleotide base residue sequence from AAA to CAG which resulted in an amino acid residue change from lysine to glutamine as shown in Figure 8A and 8B.

The individual single-stranded polynucleotides (N_1, N_4, N_6 and N_7), which were modified from their counterparts listed in Table 9, are listed in Table 10 below.

Table 10

Seq. Id. No.		
(78)	N1)	5' AGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCAT 3'
(67)	N2)	5' AATGAAATACCTATTGCCTACGGCAGCCGCTGGATT 3'
(68)	N3)	5' GTTATTACTCGCTGCCCAACCAGCCATGGCCC 3'
(79)	N4)	5' AGGTGCAGCTGCTCGAGAATTCTAGACTAGGTTAATAG 3'
(70)	N5)	5' TCGACTATTAACTAGTCTAGAATTCTCGAG 3'
(80)	N6)	5' CAGCTGCACCTGGGCCATGGCTGGTTGGG 3'
(72)	N7)	5' CAGCGAGTAATAACAATCCAGCGGCTGCCGTAGGCAATAG 3'
(81)	N8)	5' GTATTTCATTATGACTGTCTCCTTGAAATAGAATTTGCGGCCGC 3'
(74)	N9-4)	5' AGGTGAAACTGCTCGAGATTTCTAGACTAGTTACCCGTAC 3'
(75)	N11)	5' GACGTTCCGGACTACGGTTCTTAATAGAATTCG 3'
(76)	N12)	5' TCGACGAATTCTATTAAGAACCGTAGTC 3'
(77)	N10-5)	5' CGGAACGTCGTACGGGTAACTAGTCTAGAAATCTCGAG 3'

The modified ImmunoZAP H vector was created to eliminate an unnecessary Sac I site in the ImmunoZAP H vector, (Example 9), when the heavy and light chain vectors were combined. The modifications also improved the efficiency of secretion of positively changed amino acids in the amino terminus of the expressed protein. Inouye et al., Proc. Natl. Acad. Sci. USA, 85:7685-7689 (1988).

9. Preparation of V_L Expression Vector ImmunoZAP L Construction

To express the plurality of V_L coding polynucleotides in an E. coli host cell, a vector was constructed that placed the V_L coding polynucleotide in the proper reading frame, provided a ribosome binding site as described by Shine et al., Nature, 254:34, (1975), provided a leader sequence directing the expressed protein to the periplasmic space and also provided a polynucleotide that coded for a spacer protein between the V_L polynucleotide. A synthetic DNA sequence containing all of the above polynucleotides and features was constructed by designing single stranded polynucleotide segments of 20-40 bases that would hybridize to each other and form the double stranded synthetic DNA sequence shown in Figure 6B. The individual single-stranded polynucleotides (N_1-N_8) are shown in Table 9 above.

Polynucleotides N2, N3, N4, N6, N7 and N8 were kinased by adding 1 μl of each polynucleotide and 20 units of T_4 polynucleotide kinase to a solution containing 70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DDT, 10 mM 2ME, 500 micrograms per ml of BSA. The solution was maintained at 37°C for 30 minutes and the reaction stopped by maintaining the solution at 65°C for 10 minutes. The two end polynucleotides 20 ng of polynucleotides N1 and polynucleotides N5 were added to the above kinasing reaction solution together with 1/10 volume of a solution containing 20 mM Tris-HCl, pH 7.4, 2 mM $MgCl_2$ and 50 mM NaCl. This solution was heated to 70°C for 5 minutes and allowed to cool to room temperature, approximately 25°C, over 1.5 hours in a 500 ml beaker of water. During this time period all the polynucleotides annealed to form the double stranded synthetic DNA insert. The individual

polynucleotides were covalently linked to each other to stabilize the synthetic DNA insert with adding 40 µl of the above reaction to a solution containing 50 µl Tris-HCl, pH 7.5, 7 mM $MgCl_2$, 1 mM DTT, 1 mM ATP and 10 units of T4 DNA ligase. This solution was maintained at 37°C for 30 minutes and then the T4 DNA ligase was inactivated by maintaining the solution at 65°C for 10 minutes. The end polynucleotides were kinased by mixing 52 µl of the above reaction, 4 µl of a solution recontaining 10 mM ATP and 5 units of T4 polynucleotide kinase. This solution was maintained at 37°C for 30 minutes and then the T4 polynucleotide kinase was inactivated by maintaining the solution at 65°C for 10 minutes.

The completed synthetic DNA insert was ligated directly into a Lambda Zap II vector prepared in Example 7 that had been previously digested with the restriction enzymes Not I and Xho I. The ligation mixture was packaged according to the manufacture's instructions using Gigapack II Gold packing extract and the packaged ligation mixture was plated on XL1-Blue cells as described in Example 8a. Individual lambda Zap II plaques were cored and the inserts excised according to the in vivo excision protocol as described in Example 8a. This in vivo excision protocol converts the cloned insert from the Lambda Zap II vector into a phagemid vector to allow easy manipulation and sequencing and also produces the phagemid version of the V_L expression vectors. The accuracy of the above cloning steps was confirmed by sequencing the insert using the Sanger dideoxy method described by Sanger et al., Proc. Natl. Acad. Sci. USA, 74:5463-5467, (1977) and using the manufacturer's instructions in the AMV reverse transcriptase ^{35}S-dATP sequencing kit (Stratagene). The sequence of the

PCT/US92/01475

resulting V_L expression vector is shown in Figure 6B and Figure 9.

The V_L expression vector used to construct the V_L library was the phagemid produced to allow the DNA of the V_L expression vector to be determined. The phagemid was produced, as detailed above, by the in vivo excision process from the Lambda Zap V_L expression vector (Figure 9).

10. Construction of V_{HL} Expression Vectors and Library

a. Ligation of Dicistronic DNA Molecules with Modified ImmunoZAP H

In preparation for cloning a library enriched in V_H-V_L-coding (V_{HL}) dicistronic DNA molecules, PCR amplified products (human or mouse) prepared in Examples 4, 5 and 6 (50 mM NaCl, 25 mM Tris-HCl, pH 7.7, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 100 ug/ml BSA, at 37°C were digested with restriction enzymes Xho I and Xba I at a concentration of 60 units of enzyme per ug of DNA, and purified on a 1% agarose gel. After gel electrophoresis of the digested PCR amplified dicistronic DNA molecules, the region of the gel containing the DNA fragments of approximately 1360 base pairs in size was excised, purified using Gene-Clean (BIO 101, La Jolla, California), ethanol precipitated and resuspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA to a final concentration of 10 ng/ul. Equimolar amounts of the insert were then ligated overnight at 4°C to 1 ug of modified ImmunoZAP H vector, prepared in Example 8b, (Stratagene) previously digested with Xho I and Xba I. A portion of the ligation mixture (1 ul) was packaged for 2 hours at room temperature using Gigapack Gold

packaging extract (Stratagene) and the packaged material was plated on a permissive E. coli (strain XL1-blue) lawn to generate plaques. The library was determined to consist of predominantly V_{HL} with less than 5% non-recombinant background.

b. Screening of Antibody-Producing Plaques

1) Human

To screen for expression of V_{HL} dicistronic molecules, E. coli were infected to yield approximately 100 plaques per plate. Replica filter lifts of the plaques on an agar plate were produced by overlaying a nitrocellulose filter that had been soaked in 10 mM isopropyl beta-dithiogalactopyranoside on each plate with transfer for 15 hours at 23°C. For detection of V_{HL} antibody fragment expression, the filters were screened with rabbit anti-human heavy and light chain antibodies followed by goat anti-rabbit antibody coupled to alkaline phosphatase (Cappel Laboratories, Malvern, Pennsylvania). The detection of immunoreactive product confirmed the presence and expression of V_{HL} antibody fragments.

To identify human DNA clones expressing antibody that bound TT, plaques were plated and proteins expressed as described above. Replica filters were incubated with 0.2 nM ^{125}I-tetanus toxoid and washed. Positive plaques were identified by autoradiography and isolated. The frequency of positive clones in the library was equivalent to (number of positive clones)/[number of plaques screened) X (fraction of plaques expressing V_{HL}). Concentrated nonadsorbed tetanus toxoid was iodinated with sodium iodide ^{125}I (ICN, Irvine, California) by the Chloramine-T method as described in Botton et al., Biochem. I., 133:529-539 (1973) and available in a kit (Iodo-Beads, Pierce, Rockford, Illinois).

Human DNA clones were re-plated at approximately 100 phage per plaque side by side with the parental phage that were used as templates for PCR amplification and screened in the primary antigen binding screen. The results of the screening procedure are seen in Figure 11. Similar signals between the parental clones and the V_{HL} dicistronic DNA molecules demonstrated that the sequence differences introduced with the C_H1' and V_L primers did not adversely affect gene expression. Also, it should be noted in Figure 11 that a random parental clone that did not react with tetanus toxoid, 7G1, was unreactive before and after the PCR dicistronic fusion, as was the control ImmunoZAP H vector (IZ H).

2) Mouse

Mouse antibody-producing plaques prepared in Example 7 were screened for antibody expression with rabbit anti-mouse heavy and light chain antibody (Cappel Laboratories) as described above.

11. Characterization of Cloned Dicistronic V_{HL} Repertoire in Expression Library

a. Verification of Presence and Size of Cloned Dicistronic V_{HL} Repertoire

Bacteriophage from purified reactive plaques prepared in Example 10b were converted to the plasmid format by in vivo excision with R408 helper phage according to manufacturer's protocol (Stratagene) and also described in Short et al., Nucl. Acids Res., 16:7583-7600 (1988). In the in vivo excision protocol, the cloned insert from the ImmunoZAP H vector was converted into a phagemid vector to allow easy manipulation and sequencing. Briefly, phage plaques were cored from the agar plates and transferred to sterile microfuge tubes containing

500 ul of a buffer containing 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% (w/v) gelatin and 20 ul of chloroform.

For excisions, 200 ul of the phage stock, 200 ul of XL1-Blue cells (A_{600} = 1.00) and 1 ul of R408 helper phage (1 x 10^{10} plaque forming units (pfu)/ml) were incubated at 37°C for 15 minutes. After a 4 hour incubation in Luria-Bertani (LB) broth and heating at 70°C for 20 minutes to heat kill thc XL1-blue cells, the phagemids were re-infected into XL1-Blue cells and plated onto LB plates containing ampicillin. Double stranded DNA was prepared from the phagemid containing cells according to the methods described by Holmes et al., Anal. Biochem., 114:193, (1981). Clones were first screened for DNA inserts by restriction digests with Xho 1 and Xba 1. The detection of 1390 base pair fragment on an agarose gel confirmed the presence of a V_{HL} dicistronic molecule insert.

b. Sequencing of Plasmids from Expression Library

Clones containing the putative V_{HL} insert were sequenced using reverse transcriptase according to the general method described by Sanger et al., Proc. Natl. Acad. Sci., USA, 74:5463-5467, (1977) and the specific modifications of this method provided in the manufacturer's instructions in the AMV reverse transcriptase ^{35}S-dATP sequencing kit (Stratagene).

Nucleotide sequence analysis of several fusion clones indicated that the sequence of the fusion region was identical to that shown in Figure 5, proving that the clones were actually generated through a fusion PCR intermediate.

c. Advantages of Fusion-PCR to Produce Dicistronic DNA Molecules

PCR amplification can, therefore, be used to fuse sequences responsible for encoding subunits of a heterodimeric protein together into a single DNA fragment that can then direct the expression of both subunits from one expression vector. In the case of antibodies, if the source of nucleic acid template comes from hybridoma mRNA, there is only one heavy and light chain sequence to choose from, and thus the heavy:light pair is a "natural" pair.

However, if spleen, peripheral blood B-cell, or other lymphocyte mRNA is used as the source of template, the PCR fusion reaction to form a dicistronic DNA molecule can randomly pair heavy and light chains from different cells, producing a combinatorial library. In such a library, only a small fraction of the clones contain the original heavy and light chain pairs. This may not be a problem if the desired natural pair is well represented in the original B-cell population, as is the case with hyperimmunized donors. However, if one wishes to find a naturally occurring rare specificity in a combinatorial library, one may have to screen an large number of clones.

The fusion method presented here may offer a solution to the random combinatorial problem. If one begins with a very dilute population of B-cells (possibly in a medium that limits diffusion), it may be possible for the dicistronic event to occur between naturally paired heavy and light chain sequences before significant mixing between B-cell RNA occurs. Thus, the fused heavy and light chain sequences would be the original pairs, and the resulting library would express predominantly the naturally occurring antibody specificities. Such a library would be highly preferable when rare natural specificities are sought.

Another advantage to this method is that only one vector and one cloning step are necessary. This saves a substantial amount of time, resources, and effort. Moreover, the ease of the single PCR reaction greatly simplified the process of going from B-cell RNA to an E. coli library, making this approach a noteworthy alternative to standard hybridoma technology.

12. Production of An Expression Vector for Fusing the LamB Outer Membrane Spanning Signal to a Polypeptide

The following PCR primers are used to produce a DNA segment encoding the surface expression signal amino acid residue sequence of lamB, (i.e., residue positions 51-184 as shown in Figure 3):

Table 11
LamB Primers

Seq. Id. No.		
(82)	upstream[1]	5' TTACTAACTAGTTTCTATTTCGACACTAACGTG3'
(83)	downstream[2]	5' TTAGATCTAGATTTCCATCTGCGCTAAACGCAC3'

[1] Underlined sequence designates the location of an Spe I restriction site.

[2] Underlined sequence designates the location of an Xba I restriction site.

The primers are mixed pairwise with genomic DNA used from E. coli having the lamB gene as template. The amplified DNA segment is purified by preparative agarose gel electrophoresis, digested with Spe I and Xba I restriction endonucleases, and

subsequently ligated into ImmunoZap H/L at the Spe I restriction sites flanking the decapeptide tag sequence, i.e., the decapeptide tag sequence is replaced with the outer membrane spanning signal sequence.

The immunoZAP vector (H/L) is created from the heavy and light chain libraries, prepared in Examples 8 and 9, respectively, by fusing the vectors at the Eco R1 site as follows. DNA is purified from the light chain library and restriction digested with Mlu 1 and Eco R1. This cleaves the DNA from the left arm of the vector into several pieces while leaving the right arm with the light chain inserts intact. DNA is purified from the heavy chain libraries and restriction digested with Hind III and Eco R1. This cleaves the DNA from the right arm of the vector into several pieces while leaving the left arm with the heavy chain inserts intact. The intact left arm of the heavy chain vector containing the heavy chain inserts and right arm of the light chain vector containing the light chain inserts are then mixed and ligated at the common Eco R1 restriction site. The resultant ImmunoZAP H/L vector is shown in Figure 12. The ligations and packaging are as described in Example 2 to create the ImmunoZAP H/L library.

A DNA segment coding for a preselected polypeptide, such as a V_H, can then be ligated into the lamB-modified ImmunoZap H expression vector at position between, and is the same reading frame with, the pelB leader and the lamB signal sequences. The vector thus produced expresses the preselected polypeptide as a double-fusion protein, i.e., having pelB leader and lamB surface expression signal polypeptide segments operatively linked to the preselected polypeptides amino- and carboxy-termini,

respectively.

To increase the distance of the preselected polypeptide from the surface membrane of E. coli which results in decreased steric hinderance and competition of the preselected polypeptide with the lipopolysaccharide coat of E. coli, inserts of various lengths can be constructed in the recombinant plasmid vector (Tables 12 and 13).

The amino acid residue sequence of the inserts A, B, C or D following expression in E. coli as described in Example 10 are listed in Table 12.

Table 12

Insert Amino Acid Residue Sequence

Seq. Id No.	
(84)[1]	GluProLysSerCysAspLysThrHisThrSerProProAla ProAlaProGluLeuLeuLysSerSerPheTyrPheAspThr
(85)[2]	ProLysSerCysAspLysThrHisThrGluProLysSerThr AspLysThrHisThrSerProProAlaProAlaProGluLeu LeuLysSerSerPheTyr
(86)[3]	ProLysSerCysAspLysThrHisThrSerLysSerSerPhe TyrPheAsp
(87)[4]	GluProLysSerCysAspLysThrHisThrSerTyrPheTyr AspValProAspTyrGlySerLysSerSerPheTyrPheAsp Thr

[1] Insert A: Moves Spe 1 site, retains native IgG1 upper hinge region; retains original lamB sequence.

[2] Insert B: Moves Spe 1 site, retains original IgG1 upper hinge region, retains original lamB sequence.

[3] Insert C: Moves Spe 1 site, retains original IgG1 upper hinge region, duplicates 10 amino acids in the upper hinge region.

[4] Insert D: Moves Spe 1 site, retains native IgG1

upper hinge region; retains original lamB sequence.

Table 13

Insert Primers

Seq Id No.	
(88)[1]	5' GTCCTGTCCGAGGTGCAGCTGCTCGAGTCTGG 3'
(89)[2]	5' GTCCACCGGCCCCAGCACCTGAACTCCTGAAGAGCAGC TTCTAT 3'
(90)[3]	3' CTCGGGTTTAGAACACTGTTTTGAGTGTGATCAGGTGGC CGGGGTCGTGGAC 5'
(89)[4]	5' GTCCACCGGCCCCAGCACCTGAACTCCTGAAGAGCAGC TTCTAT 3'
(90)[5]	3' CTCGGGTTTAGAACACTGTTTTGAGTGTGATCAGGTGGC CGGGGTCGTGGAC 5'
(91)[6]	5' GTGACAAAACTCACACTAGTAAGAGCAGCTTCTAT 3'
(92)[7]	3' CTCGGGTTTAGAACACTGTTTTGAGTGTGATCA 5'
(93)[8]	5' GTGACAAAACTCACACTAGTTACCCGTACGACGTTCCGGAC TACGGTTCTAAGAGCAGCTTGTAT 3'
(94)[9]	3' CTCGGGTTTAGAACACTGTTTTGAGTGTGATCATCA 5'
(95)[10]	3' CACGCAAATCGCGTCTACCTTAAGATCGTTCCTCTGTCA GTATTACTTTATGGATAACGGATGCCGTCGGCGACCTAA CAATAA 5'
(96)[11]	5' GCCTACGGCAGCCGCTGGATTGTTATTAATCGCTGCCCA ACCTGCCATGGCTGAGCTCGTGATGACCCATGCTCC 3'
(97)[12]	5' TCCTTCTAGATTACTAACACTCTCCCCTGTTGAAGCTCT TTGTGACGGGCGAACTC 3'

[1] 5' heavy chain V_H primer used for all constructions.

[2] 5' lamB overlapping primer for insert A.

[3] 3' heavy chain C_H1 overlapping primer for insert A.

[4] 5' lamB overlapping primer for insert B.

[5] 3' heavy chain C_H1 overlapping primer for insert B.

[6] 5' lamB overlapping primer for insert C.

[7] 3' heavy chain C_H1 overlapping primer for insert C.
[8] 5' lamB overlapping primer for insert D.
[9] 3' heavy chain C_H1 overlapping primer for insert D.
[10] 3' lamB overlapping primer with 5' light chain primer.
[11] 5' light chain overlapping primer with 3' light chain prime.
[12] 3' lamB overlapping primer with 5' light chain primer.

The inserts between the heavy chain and lamB sequences are made using the PCR-fusion procedure for producing dicistronic DNA as prepared in Examples 2 and 3 with the following exceptions. The light chain and lamB sequences are fused together using the outside primers and limiting amounts of the inside primers (Table 13). The resultant PCR products are gel purified using Gene Clean (BIO 101) as described in Example 10 before PCR-fusing it to the heavy chain using only outside primers (Table 13). The resultant PCR-fusion product consists of V_H-insert A, B, C or D-lamB-light chain. The region inserted by the PCR primers between the lamB and light chain creates the same dicistronic bridge previously inserted between the heavy and light chain DNAs. This product is ligated with the modified ImmunoZAP H vector restriction digested with the enzymes Xho I and Xba I as prepared in Example 10. After insertion, the dicistronic message encoded by the DNA allows expression of the heavy chain and lamB as a fusion protein and the light chain as a separate protein.

Surface expression is accomplished by transforming the recombinant plasmid vector into an E. coli strain, lacking its endogeneous lamB gene, thus avoiding competition between the recombinant lamB

signal-containing product and a normal lamB gene product for available membrane-spanning sites. A preferred host is a lamB deletion mutant of the E. coli SURE strain. The E. coli SURE strain is commercially available from Stratagene.

13. Production of Antigen-Specific B Cells

a. In Vitro Immunization

1) Preparation of T Cell Replacing Factor, s-PWM-T

Blood was collected from healthy donors and PBLs were isolated as described in Example 2. Isolated PBLs were then fractionated into T and non-T cells by AET-SRBC (2-aminoethylthiouronium bromide-sheep red blood cell) rosetting according to the procedure described by Callard. Callard et al., Eur. J. Immunol., 11, 206 (1981). Briefly, the isolated PBLs were treated with a 1% suspension of AET-modified sheep red blood cells. The rosette was purified over a Ficoll gradient and the red blood cells removed by hypotonic lysis.

The procedure for preparing the T cell replacing factor, s-PWM-T, was performed as described by Danielson. Danielson et al., Immunol., 61:51-55 (1987). The resultant enriched T cell population was suspended in RPMI-1640 medium (Gibco Laboratories, Santa Clara, California) supplemented with 1% (v/v) non-essential amino acids, 4 mM L-glutamine, streptomycin (50 ug/ml), penicillin (50 IU/ml) and 10% heat-inactivated human AB serum at a concentration of 2×10^6 cells/ml, and irradiated (2000 rads; 1 rad = 0.01 Gy). Following irradiation, the T cells were activated by treatment with 10 ug of pokeweed mitogen (PWM)/ml (Sigma) for 24 hours at 37°C. The supernatant was collected and stored at 4°C. PWM

activation of T cells results in secretion of gamma interferon, interleukin-2 (IL-2) and various undefined B cell growth factors into the medium. Growth factor containing supernatant from the PWM-treated T cells, hereinafter designated s-PWM-T, was collected and added to lymphocyte cell cultures prepared below.

2) Preparation of In Vitro Immunization Cultures

The procedure for in vitro immunization of PBLs was performed as described by Borrebaeck. Borrebaeck et al., Proc. Natl. Acad. Sci. USA, 85:3995-3999 (1988). Human PBLs, isolated as described in Example 2, were resuspended to a concentration of 1 X 10^7 cells/ml in serum-free, supplemented RPMI-1640, prepared as described above, containing 2.5 mM L-leucine methyl ester hydrochloride (Leu-OMe) from a 0.5 M stock solution prepared in water. (Sigma Chemical Co., St. Louis, Missouri). The cells were incubated at room temperature for 40 minutes and then washed three times in RPMI-1640 containing 2% heat-inactivated human serum. Cell recovery after treatment with Leu-OMe ranged from 30-90%. The treatment with Leu-OMe was performed to effect the removal of a Leu-OMe-sensitive subpopulation leaving a population of cells that respond to T-cell dependent antigen stimulation in vitro.

Leu-OMe-treated PBLs were immunized in vitro with either keyhole limpet hemocyanin (KLH) (Sigma) or tetanus toxoid (TT) (Example 2). For this protocol, and for the subsequent ELISPOT assays, the Leu-OMe-treated T cells were first suspended in supplemented RPMI-1640, containing 50 uM beta-mercaptoethanol, 10% heat-inactivated human AB serum, 30% (v/v) s-PWM-T, and antigen (1-1000ng/ml). For determination of total

immunoglobulin content in the culture supernatants, the cells were maintained in heat-inactivated fetal bovine serum instead of human AB serum. The antigen-treated Leu-OMe-treated PBLs were then plated at a concentration of 2 X 10^6 cells/ml in a 4-ml (six-well plates) or 30-ml (75-cm^2 flask) and maintained at 37°C in 5% CO_2 for three days. The cells were pelleted and washed one time with RPMI-1640 supplemented medium prepared above lacking antigen to effect the removal of antigen. The washed antigen-treated cells were resuspended in fresh medium containing s-PWM-T, but lacking antigen. The cells were thereafter cultured for three to four more days for a total maintenance period of six to seven days, at which time the levels of antigen-specific antibody and/or the number of antigen-specific antibody secreting cells were determined by ELISA and ELISPOT assays, respectively.

b. Immunoassays

1) ELISA Assay for Determining the Levels of Antigen-Specific Antibody

The antigen-specific immunoglobulin (IgM and IgG) secreted into the medium from antigen-treated PBLs prepared above was determined by ELISA. Briefly, 100 ng/ml of antigen, either KLH or TT, diluted in PBS, pH 7.5, was added to individual wells of 96-well microtitre plates. The plates were allowed to stand at room temperature for 16 to 18 hours. After removing unabsorbed proteins, wells were blocked with 0.2% gelatin-PBS for 1 hour at room temperature.

One hundred ul of culture medium samples were added to the antigen-coated wells and incubated at room temperature for 1 hour. The wells were then rinsed three times with PBS-containing 0.05% Tween 20. Alkaline phosphatase (AP) conjugated to isotype-specific antibodies (goat-anti-human IgM or IgG) (1

ug/ml-final concentration) (Boehringer Mannheim, Indianapolis, Indiana) was diluted in 50 mM PBS, pH 7.5, containing 1.5 M sodium chloride and 0.1% Tween 20 and 100 ul of diluted AP antiglobulin conjugate were then added to each well and maintained at room temperature for one hour or at 4°C overnight. The wells were then rinsed three times with PBS 0.05% Tween 20.

The washed wells were then inverted to remove the remaining buffer. A 1 mg/ml solution of PNPP (Sigma) p-nitro phenylphosphate dissolved in PBS was then added to each well for detection of antigen-specific antibodies and optical density measured at 405 nm.

2) ELISPOT Assay for Determining the Number of Antigen-Specific Antibody-Secreting Cells

ELISPOT assays are performed as described by Czerkinsky. Czerkinsky et al., J. Immunol. Methods, 65:190-121 (1983). For measuring the number of antigen-specific antibody-secreting cells in the in vitro immunized PBL cultures ELISPOT was performed. For this assay, 3.5 centimeter diameter polystyrene petri dishes (Falcon, Oxnard, California) were filled with 1.5 ml of either KLH or TT antigen at a concentration of 1 ug/ml. Borrebaeck et al., supra. The plates were washed as described for the ELISA assay. The antigen-coated plates were then blocked with 0.2% gelatin at 37°C. Lymphocytes (10^5 to 10^6) were added to each dish and allowed to incubate undisturbed overnight at 37°C. The cells were removed and the plates were washed twice with cold PBS and then maintained for 10 minutes with cold 10 mM EDTA-PBS. The plates were then rinsed three times with PBS containing 0.5% Tween-20. Antigen-specific human

antibody was detected with alkaline phosphatase-goat anti-human IgG or IgM, followed by the addition of enzyme substrates 150 ug/ml BCIP (5-bromo-4-chloro-3-indolyl phosphate), and 300 ug/ml NBT (nitroblue tetrazolium), dissolved in molten 1% agarose in PBS. The plates were then incubated for one to several hours at room temperature, after which time, blue spots appeared corresponding to the positions of antigen-specific antibody secreting cells.

The frequency of antigen-specific B cells was determined as (number of antigen-specific antibody secreting cells)/(number of B cells added per plate). In vitro immunization was demonstrated by an increase in the frequency of antigen-specific B cells, in response to antigen stimulation. The number of B cells added to each ELISPOT plate was assumed to be approximately 10% of the total number of Leu-OMe-treated PBLs based on immunofluorescence analysis of Ohlin. Ohlin et al., Immunology, 66:485-90 (1989). The total number of lymphocytes was determined by trypan exclusion. The results of these assays are described below.

3) Panning to Increase the Frequency of Antigen-Specific B Cells

PBLs, prepared in Example 2, were treated with Leu-OMe and resuspended in supplemented RPMI-1640 medium containing 2% fetal bovine serum as described above. Approximately 1 to 10 x 10^6 Leu-OMe-treated PBLs were added to polystyrene petri dishes, previously coated with 1 µg/ml of either KLH or TT antigen and blocked with 0.2% gelatin. The cells were then maintained at 4°C for 1 hour. After the non-adherant cells were decanted, the plates were washed three times with chilled medium and the non-adherent fractions were pooled.

Depletion of antigen-specific B cells was demonstrated by culturing the non-adherent cells in the presence of s-PWM-T, as described above, for 6 days. The number of antigen-specific antibody producing cells was then determined by the ELISPOT assay. The number of B cells which adhere under the conditions described above was determined using two different methods. An enriched population of B cells was obtained by rosetting with AET-treated sheep red blood cells. The non-rosetting cells were then panned on autologous plasma-coated petri dishes, and the non-adherent lymphocytes (B cells) recovered. In one set of experiments, the B cells were labelled overnight with ^{35}S-methionine, panned as described above, and the percent radioactivity adhering to the dishes determined. In the second set of experiments, the number of purified cells which adhered was determined microscopically using an ocular grid. The results of the experiments are described below.

4) Panning In Vitro Immunized Cells

Panning and in vitro immunization were combined to enrich the frequency of antigen-specific B cells beyond the level which can be achieved with either technique alone. In vitro stimulated lymphocytes were cultured as described above for 5 days, transferred to fresh medium and panned as described on antigen coated dishes. The results of these experiments are described below.

c. Results of In Vitro Immunization and Cell Panning

Using KLH and tetanus toxoid (TT) as model antigens in the above-described procedure resulted in a 2-3 fold increase in the frequency of both TT- and KLH-specific B cells. The frequency of KLH-specific B cells was considerably influenced by

resuspending the cells in fresh media, containing T cell supernatant and lacking antigen, several days after exposure to antigen. As Table 14 shows, a 7 to 17-fold increase in the number of antigen-specific B cells was observed when the cells were pulsed with antigen for 2 to 3 days. Exposing the cells for the entire culture period (6-7 days), on the other hand, resulted in an average increase of only 3-fold.

Table 14

Expansion of KLH-specific B cells*

Expt.	Donor	Day Media Exchange/KLH Removed						
		1	2	3	4	5	6	7
1	1	1	17	4	2	16	6	nd
2	1	nd	>9	1	3	4	4	nd
	2	nd	1	7	1	2	1	nd
3	2	nd	14	2	nd	1	nd	2
	3	nd	>15	>11	nd	1	nd	4

* ratio = (# anti-KLH Ig secreting cells cultured with s-PWM-T and KLH)/(# anti-KLH Ig secreting cells cultured with s-PWM-T) detected with ELISPOT

As expected for induction of a primary immune response, the KLH-specific antibodies secreted were of the IgM isotype. Antigen pulsing resulted in an average increase of 9-fold and a mean frequency of 3.3 x 10^{-3} (Table 15). These results indicate that primary immunization of naive human B cells can give rise to frequencies of antigen-specific B cells which are comparable to those found when B cells were collected from human donors boosted with TT. Mullinax

et al., supra.

Table 15
Frequency of KLH-specific Cells
After Primary Immunization

# Expts	# Donors	Antigen Pulse [Ag] (ng/ml)	Pulse Time (days)	Freq anti-KLH IgM B Cells Background (range)	+ Ag (range)
4	6	400-800	2-3	$3.8X10^{-4}$ ($0.6-7.9X10^{-4}$	$3.3X10^{-3}$ ($0.9-8.7X10^{-3}$)

In addition to in vitro immunization, cell panning techniques have also been developed for the enrichment of antigen-specific B cells. Table 16 summarizes the degree of enrichment observed for a single cycle of panning against a model antigen.

Table 16
Enrichment of Antigen-Specific B Cells

Expt. #	% Anti-TT Ig Secreting Cells Adhered	Expt. #	% B Cells Adhered	Enrichment Factor
1	100% ± 0%	3	1.5% ± 0.6%	67
2	90% ± 14%	4	10% ± 1.9%	9

Peripheral blood lymphocytes from unboosted donors were panned on TT- and gelatin-coated petri dishes and the number of TT-specific B cells in the non-adherent cell population determined. In experiments 1 and 2, 100% and 90% of the anti-TT antibody secreting cells, respectively, were depleted when panned on TT plates, while only 28% and 8% were depleted when panned on gelatin (not shown). The number of purified B cells which adhere under analogous

conditions ranged from 1.5 to 10% and was determined either by labeling the cells with ^{35}S-methionine (expt. 3) or by examining the adherent cells microscopically with an ocular grid (expt. 4). These preliminary results indicate that a single cycle of cell panning can be used to increase the frequency of antigen-specific B cells by at least 9-fold, and possibly as high as 67-fold. It should be possible to further deplete B cells which bind non-specifically or with low affinity to antigen by performing sequential isolations or by altering the epitope density of the solid matrix.

By combining the results found in Tables 15 and 16, it is evident that cell panning can be used alone or in combination with in vitro immunization to increase the frequency of antigen-specific B cells in the naive repertoire by 2 to 3 orders of magnitude. Analysis of the non-adherent cells recovered after panning, before and up to 7 days after culturing with T cell supernatant (s-PWM-T), indicates that the majority of KLH-specific antibody producing cells are depleted when panned at 0 to 5 days. As Table 17 indicates, panning at days 6 and 7 (peak of antibody production) is inefficient, possibly due to either down-modulation of surface IgM receptors or interference by secreted anti-KLH antibody. To recover the greatest enrichment antigen-specific B cells, panning should be performed at day 5 to ensure maximal clonal expansion.

Table 17

# of Anti-KLH IgM-Secreting Cells in the Non-Adherent Fraction

	Panning Antigen	
Day Panned	KLH	Gelatin
2	1	9

4	1	17
5	2	17
6	17	20
7	18	5

These studies have demonstrated, with model antigens, that in vitro immunization or cell panning can be used to increase the frequency of antigen-specific B cells by at least 10-fold. Preliminary results indicate that the two techniques can be combined to give rise to frequencies which are comparable to those of the lymphocyte population used to construct the TT-specific library (10^{-3}). These techniques may obviate the requirement for in vivo immunization, thereby eliminating one of the major obstacles to the routine production of human monoclonal antibodies. By cloning human immunoglobulin sequences from E. coli expression libraries, the difficulties encountered in immortalizing antibody producing cell lines are avoided as well. Thus, preparing immunoexpression libraries from enriched populations of naive B cells should render it possible to generate human monoclonal antibodies against a variety of antigens of therapeutic and diagnostic interest.

The foregoing is intended as illustrative of the present invention but not limiting. Numerous variations and modifications can be effected without departing from the true spirit and scope of the invention.

What is claimed is:

1. A method of producing a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, said upstream and downstream cistrons encoding respective first and second polypeptides of a heterodimeric receptor, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, a repertoire of first polypeptide genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer, said inside first gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a first gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, a repertoire of second polypeptide genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer, said inside gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said first gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a second polypeptide-coding gene;

said first inside and second inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding (i) at least one stop codon in the same reading frame as said upstream cistron, (ii) a ribosome binding site

downstream from said stop codon, and (iii) a polypeptide leader having a translation initiation codon in the same reading frame as said downstream cistron, said initiation codon located downstream from said ribosome binding site;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form;

(e) separating the double stranded DNA homologs produced in steps (b) and (d);

(f) hybridizing the separated strands to form a plurality of internally-primed duplexes; and

(g) subjecting the internally-primed duplexes to conditions for primer extension to produce a plurality of different dicistronic DNA molecules, each containing a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said first polypeptide- or second polypeptide-coding DNA homologs, and said downstream cistron comprising the other of said first polypeptide- or second polypeptide-coding DNA homologs.

2. The method of claim 1 wherein steps (a)-(d) are performed concurrently in one reaction vessel.

3. The method of claim 1 wherein said signals for the initiation of translation of the downstream cistron are located downstream from the stop codon and include a ribosome binding site and a translation initiation codon encoding the first amino acid residue of a polypeptide leader, said codon located in the same reading frame as the downstream cistron.

4. A method of producing dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, said upstream and downstream cistrons encoding respective first and second polypeptides of a

heterodimeric protein, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, first polypeptide-encoding genes and a first PCR primer pair defined by an outside first gene primer and an inside first gene primer, said inside first gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of said first polypeptide gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of first polypeptide coding DNA homologs in double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, second polypeptide-encoding genes and a second PCR primer pair defined by an outside second gene primer and an inside second gene primer, said inside gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said first gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of said second polypeptide-coding genes;

said first inside and second inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding (i) at least one stop codon in the same reading frame as said upstream cistron, and (ii) signals for the initiation of translation of the downstream cistron;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of second polypeptide-coding DNA homologs in double stranded form;

(e) separating the double stranded DNA homologs produced in steps (b) and (d);

(f) hybridizing the separated strands to form a plurality of internally-primed duplexes; and

(g) subjecting the internally-primed duplexes to conditions for primer extension to produce dicistronic DNA molecules, each containing a first polypeptide-coding sequence and a second polypeptide-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said first polypeptide- or second polypeptide-coding DNA homologs, and said downstream cistron comprising the other of said first polypeptide- or second polypeptide-coding DNA homologs.

5. The method of claim 4 wherein steps (a)-(d) are performed concurrently in one reaction vessel.

6. The method of claim 4 wherein the genes of steps (a) and (c) are present in gene repertoires formed by isolating mRNA from at least 10^3 peripheral blood lymphocytes.

7. The method of claim 6 wherein said repertoire of first polypeptide genes comprises at least 10^3 different first polypeptide genes.

8. The method of claim 6 wherein said repertoire of second polypeptide genes comprises at least 10^3 different second polypeptide genes.

9. The method of claim 4 further comprising step (h) wherein said plurality of different dicistronic DNA molecules is combined with said outside first gene primer and said outside second gene primer to form a third PCR admixture, and subjecting said third PCR admixture to a plurality of PCR thermocycles to produce an amplified library of dicistronic DNA molecules.

10. The method of claim 4 wherein said outside first gene primer hybridizes to a framework, leader or promoter region of a V_H immunoglobulin gene.

11. The method of claim 4 wherein said outside second gene primer hybridizes to a J_L or framework region of a V_L immunoglobulin gene.

12. The method of claim 4 wherein said 3'-terminal priming portion of said inside first gene primer hybridizes to a J_H, hinge or framework region of a V_H immunoglobulin gene.

13. The method of claim 4 wherein said 3'-terminal priming portion of said inside second gene primer hybridizes to a framework, leader or promoter region of a V_L immunoglobulin gene.

14. A method for producing a library of dicistronic DNA molecules, which method comprises:

(a) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) a repertoire of V_H genes,

(ii) a repertoire of V_L genes,

(iii) a V_H PCR primer pair defined by an outside V_H gene primer and an inside V_H gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion; and

(iv) a V_L PCR primer pair defined by an outside V_L gene primer and an inside V_L gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to said 5'-terminal non-priming portion of said inside V_H gene primer, said V_H inside and V_L inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron,

said outside primers being present in said composition in molar excess relative to said inside

primers; and

(b) subjecting said PCR admixture to a plurality of PCR thermocycles, thereby producing said library.

15. The method of claim 14 wherein said signals include a ribosome binding site downstream from said stop codon and a translation initiation codon for said downstream cistron, said translation initiation codon located downstream from said ribosome binding site.

16. The method of claim 15 wherein said translation initiation codon is operatively linked to a polypeptide leader-encoding sequence that is in the same reading frame as the downstream cistron.

17. The method of claim 14 wherein said plurality of PCR thermocycles is at least n+5, wherein n is the number of PCR thermocycles necessary to decrease by a factor of 10 the number of said inside primers by consumption in the formation of inside primer-primed products.

18. The method of claim 17 wherein said repertoire of V_H genes and said outside V_H gene primer are present at a respective molar ratio in the range of $1:10^3$ to $1:10^7$, said repertoire of V_H genes and said inside V_H gene primer are present at a respective molar ratio in the range of $1:10^2$ to $1:10^6$, said repertoire of V_L genes and said outside V_L gene primer are present at a respective molar ratio in the range of $1:10^3$ to $1:10^7$, and said repertoire of V_L genes and said inside V_L gene primer are present at a respective molar ratio in the range of $1:10^2$ to $1:10^6$.

19. The method of claim 17 wherein said repertoire of V_H genes and said outside V_H gene primer are present at a respective molar ratio of about $1:10^4$, said repertoire of V_H genes and said inside V_H gene primer are present at a respective molar ratio of about $1:10^3$, said

repertoire of V_L genes and said outside V_L gene primer are present at a respective molar ratio of about $1:10^4$, and said repertoire of V_L genes and said inside V_L gene primer are present at a respective molar ratio of about $1:10^3$.

20. A method of producing a library of dicistronic DNA molecules comprising an upstream cistron and a downstream cistron, which method comprises:

(a) forming a polymerase chain reaction (PCR) admixture by combining, in a PCR buffer:

(i) a repertoire of V_H genes,

(ii) a repertoire of V_L genes,

(iii) an outside V_H gene primer

(iv) an outside V_L gene primer, and

(v) a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer-template portion, said 3'-terminal priming portion having a nucleotide base sequence complementary to a portion of the primer extension product of one of said outside primers, said 5'-terminal primer template portion having a nucleotide base sequence homologous to a portion of the primer extension product of the other of said outside primers and said cistronic bridge coding portion encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron; and

(b) subjecting said PCR admixture to a plurality of PCR thermocycles, thereby producing said library.

21. A method of producing an isolated dicistronic expression vector capable of expressing V_H and V_L polypeptides from respective V_H- and V_L-coding DNA homologs, said V_H and V_L polypeptide being capable of forming an antibody molecule that binds a preselected antigen, which method comprises:

(a) forming a first polymerase chain reaction (PCR) admixture by combining, in a PCR buffer, a repertoire of V_H genes and a first PCR primer pair defined by an outside V_H gene primer and an inside V_H gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a V_H gene;

(b) subjecting said first PCR admixture to a plurality of PCR thermocycles to produce a plurality of V_H-coding, double stranded DNA homologs in a double stranded form;

(c) forming a second PCR admixture by combining, in a PCR buffer, a repertoire of V_L genes and a second PCR primer pair defined by an outside V_L gene primer and an inside V_L gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said V_H gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a V_L gene, said V_H inside and V_L inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron and signals necessary for the initiation of translation of the downstream cistron;

(d) subjecting said second PCR admixture to a plurality of PCR thermocycles to produce a plurality of V_L-coding DNA homologs in double stranded form;

(e) separating said double stranded DNA homologs of steps (b) and (d);

(f) hybridizing said separated strands to form a plurality of internally-primed duplexes;

(g) subjecting said internally-primed duplexes to conditions for primer extension to produce a plurality of different dicistronic DNA molecules, each containing a V_H-coding sequence and a V_L-coding sequence linked by said cistronic bridge, said upstream cistron comprising one of said V_H- or V_L-coding DNA homologs, and said downstream cistron comprising the other of said V_H-V_L-coding DNA homologs;

(h) operatively linking for expression each of a plurality of said different dicistronic DNA molecules to expression vectors thereby forming a plurality of different V_{HL} expression vectors;

(i) transforming a population of host cells compatible with said expression vector with a plurality of said different V_{HL}-expression vectors to produce a transformed population of host cells whose members contain said V_{HL}-expression vectors;

(j) culturing said transformed population under conditions for expressing the V_H and V_L polypeptides coded for by said dicistronic DNA molecules;

(k) assaying the members of said transformed population for expression of an antibody molecule capable of binding said preselected ligand, thereby identifying transformants containing said dicistronic DNA molecule; and

(l) segregating an identified transformant to step (d) from said population, thereby producing said isolated dicistronic DNA molecule.

22. A kit comprising an enclosure containing, in separate containers, an outside first polypeptide-encoding gene primer, an outside second polypeptide-encoding gene primer, and a polynucleotide strand having a 3'-terminal priming portion, a cistronic bridge coding portion, and a 5'-terminal primer template portion, said 3'-terminal priming portion having a nucleotide base sequence

homologous to a portion of the primer extension product of one of said outside primers, said 5'-terminal primer template portion encoding a nucleotide base sequence homologous to a portion of the primer extension product of the other of said outside primers and said cistronic bridge coding portion encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron.

23. The kit of claim 22 wherein said outside first and outside second polypeptide-encoding gene primers are V_H and V_L gene primers, respectively.

24. A kit comprising an enclosure containing, in separate containers, an outside first polypeptide-encoding gene primer, an outside second polypeptide-encoding gene primer, an inside first polypeptide-encoding gene primer having a 3'-terminal priming portion and a 5'-terminal non-priming portion, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a first polypeptide-encoding gene, and an inside second polypeptide-encoding gene primer having a 3'-terminal priming portion and a 5'-terminal hybridizing portion complementary to the 5'-terminal non-priming portion of said inside first polypeptide-encoding gene primer, said 3'-terminal priming portion comprising a nucleotide sequence homologous to a conserved portion of a second polypeptide-encoding gene, said first polypeptide inside and second polypeptide inside primers, when hybridized, forming a duplex encoding a double-stranded cistronic bridge for linking said upstream and downstream cistrons, one strand of said bridge encoding at least one stop codon in the same reading frame as said upstream cistron, and signals for the initiation of translation of the downstream cistron.

25. The kit of claim 24 wherein said outside first and outside second polypeptide-encoding gene primers are V_H and V_L gene primers, respectively.

26. A library of dicistronic DNA molecules produced by the method of any one of Claims 1-3, or 14-20.

27. A dicistronic DNA molecule produced by the method of any one of Claims 4-13.

28. An isolated dicistronic expression vector produced by the method of Claim 21.

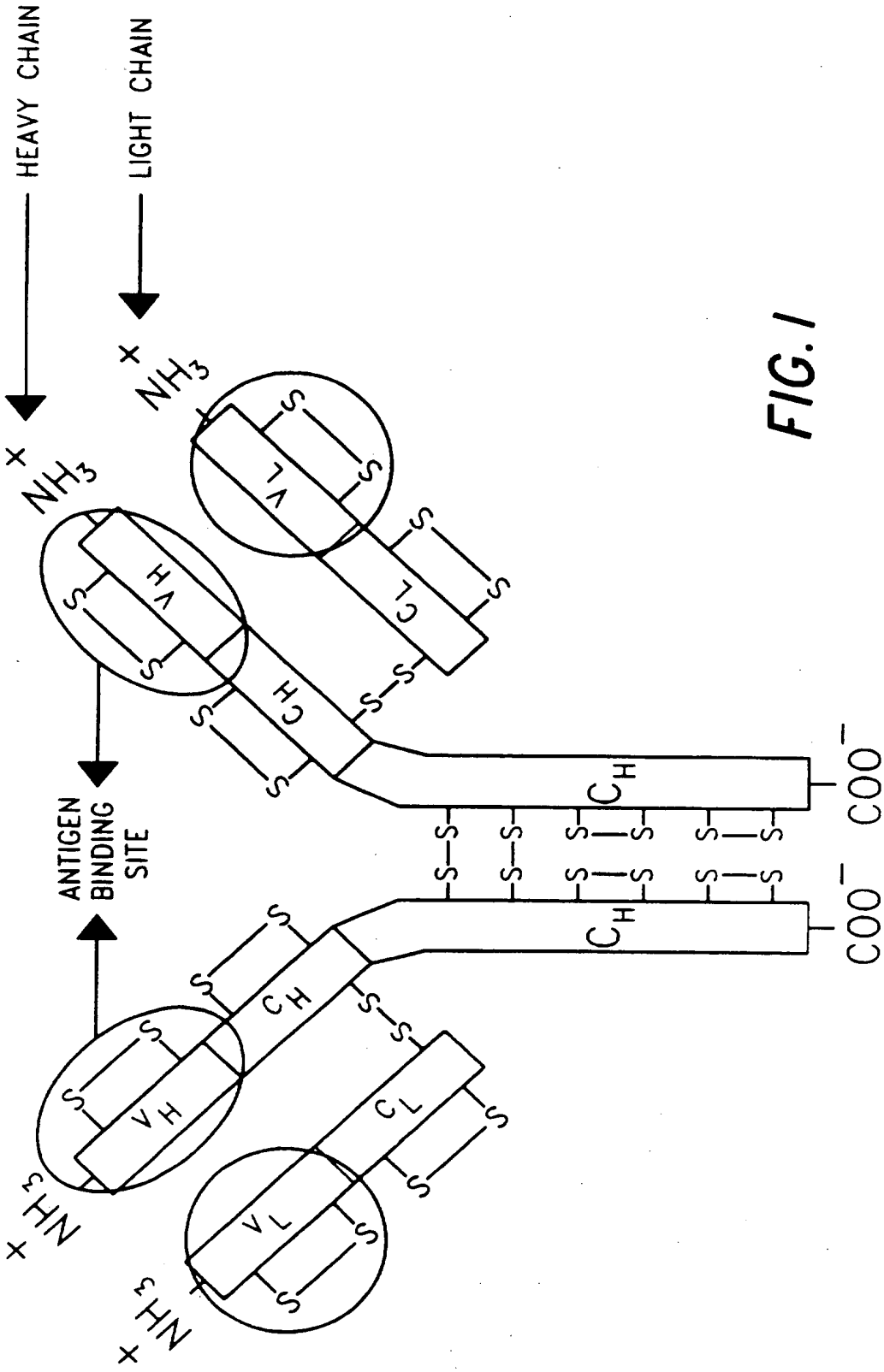

HEAVY CHAIN
LIGHT CHAIN
NH3
NH3
NH3
NH3
VL
CL
VH
CH
CH
CH
VH
CL
VL
ANTIGEN BINDING SITE
COO-
COO-
S

FIG. 1

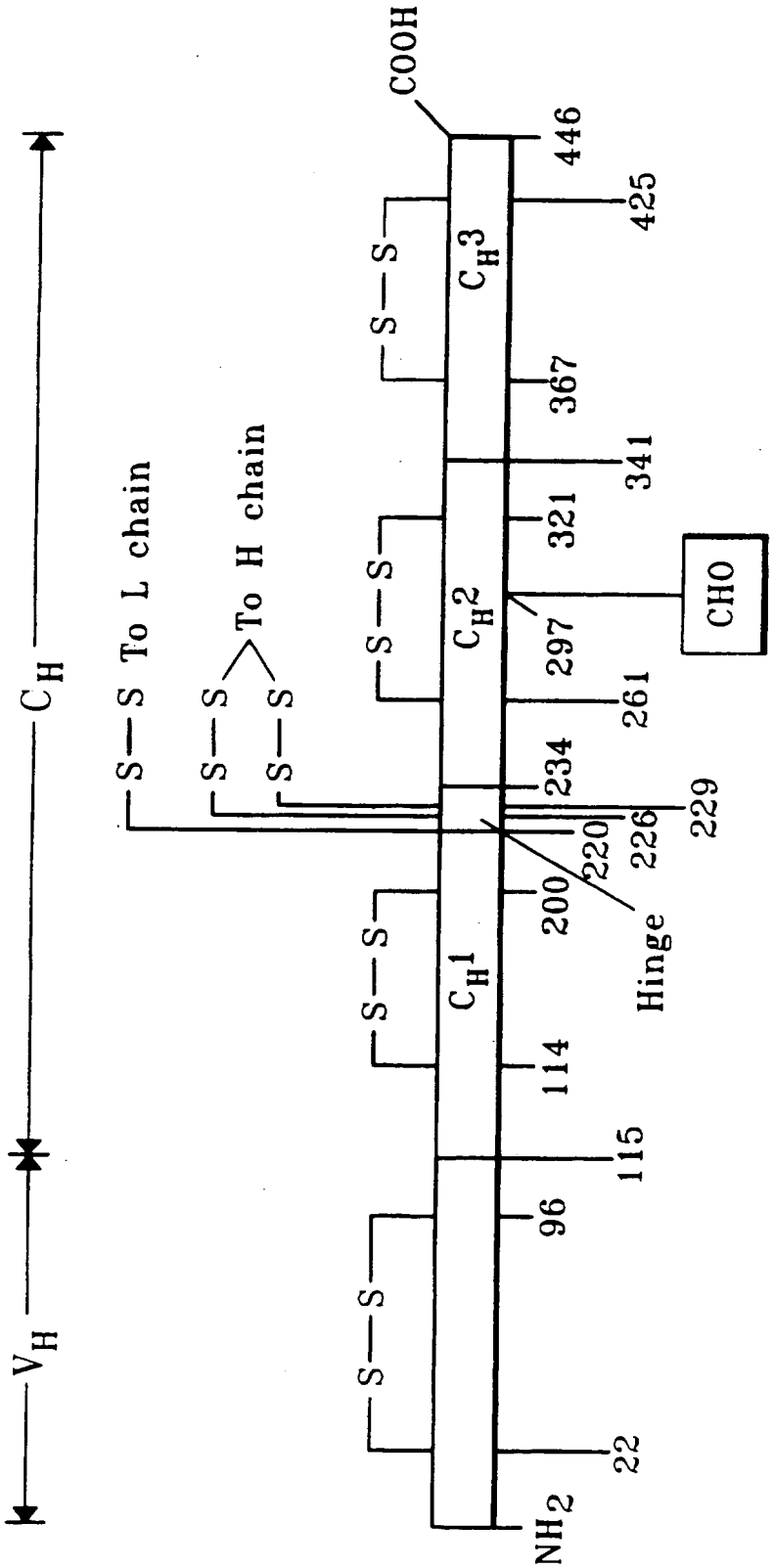

COOH
446
425
367
341
321
297
261
234
229
226
220
200
114
115
96
22
NH_2
C_H3
C_H2
C_H1
Hinge
CHO
S—S
S—S To L chain
S—S To H chain
C_H
V_H

FIG. 2A

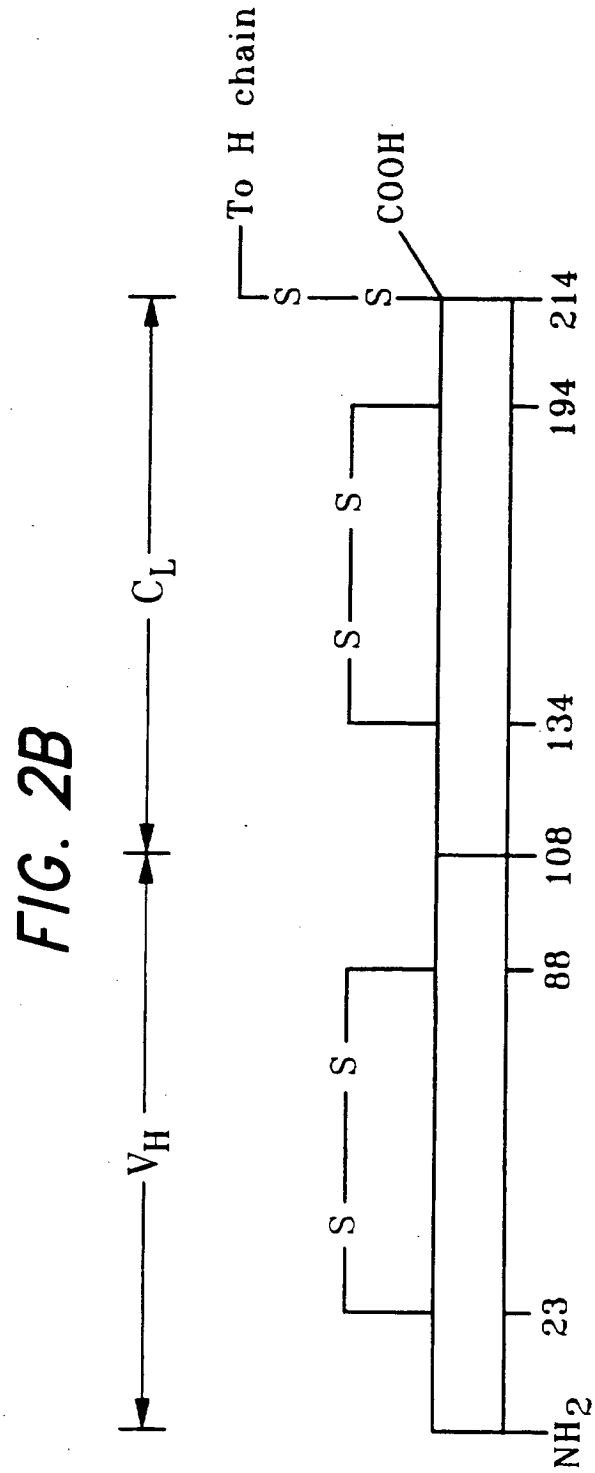

To H chain
S
S
COOH
214
194
134
108
88
23
NH_2
C_L
V_H

FIG. 2B

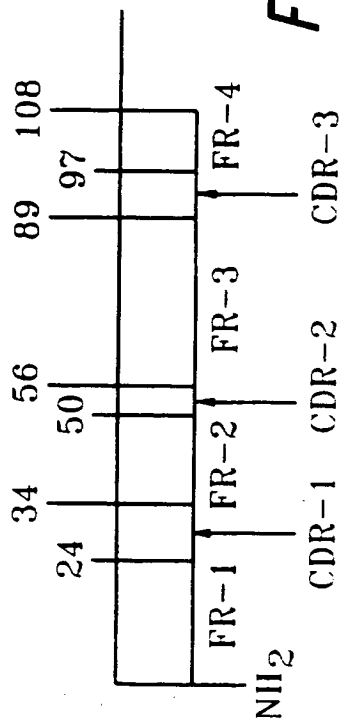

108
97
89
56
50
34
24
FR-4
FR-3
FR-2
FR-1
CDR-3
CDR-2
CDR-1
NH_2

FIG. 2C

```
250  TGGAAAGAGGGCGATAAGAGCTTCTATTTCGACACTAACGTGGCCTATTCCGTCGCACAACAG
     TrpLysGluGlyAspLysSerPheTyrPheAspThrAsnValAlaTyrSerValAlaGlnGln  71

313  AATGACTGGGAAGCTACCGATCCGGCCTTCCGTGAAGCAAACGTGCAGGGTAAAAACCTGATC
     AsnAspTrpGluAlaThrAspProAlaPheArgGluAlaAsnValGlnGlyLysAsnLeuIle  92

376  GAATGGCTGCCAGGCTCCACCATCTGGGCAGGTAAGCGCTTCTACCAACGTCATGACGTTCAT
     GluTrpLeuProGlySerThrIleTrpAlaGlyLysArgPheTyrGlnArgHisAspValHis  113

439  ATGATCGACTTCTACTACTGGGATATTTCTGGTCCTGGTGCCGGTCTGGAAAACATCGATGTT
     MetIleAspPheTyrTyrTrpAspIleSerGlyProGlyAlaGlyLeuGluAsnIleAspVal  134

502  GGCTTCGGTAAACTCTCTCTGGCAGCAACCCGCTCCTCTGAAGCTGGTGGTTCTTCCTCTTTC
     GlyPheGlyLysLeuSerLeuAlaAlaThrArgSerSerGluAlaGlyGlySerSerSerPhe  105

565  GCCAGCAACAATATTTATGACTATACCAACGAAACCGCGAACGACGTTTTCGATGTGCGTTTA
     AlaSerAsnAsnIleTyrAspTyrThrAsnGluThrAlaAsnAspValPheAspValArgleu  176

628  GCGCAGATGGAAATCAACCCGGGC
     AlaGlnMetGluIleAsnProGly                                         184
```

FIGURE 3

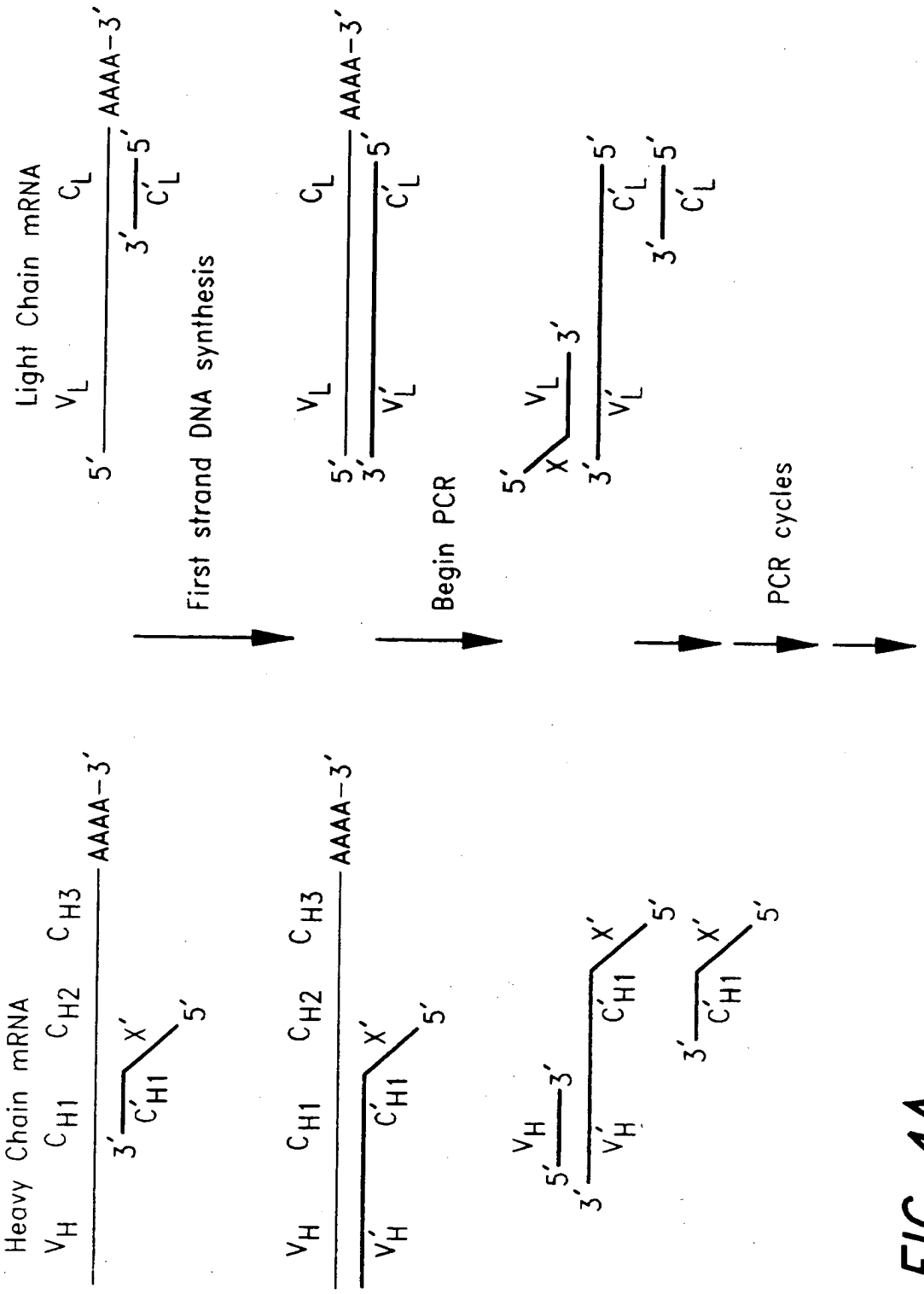


FIG. 4A

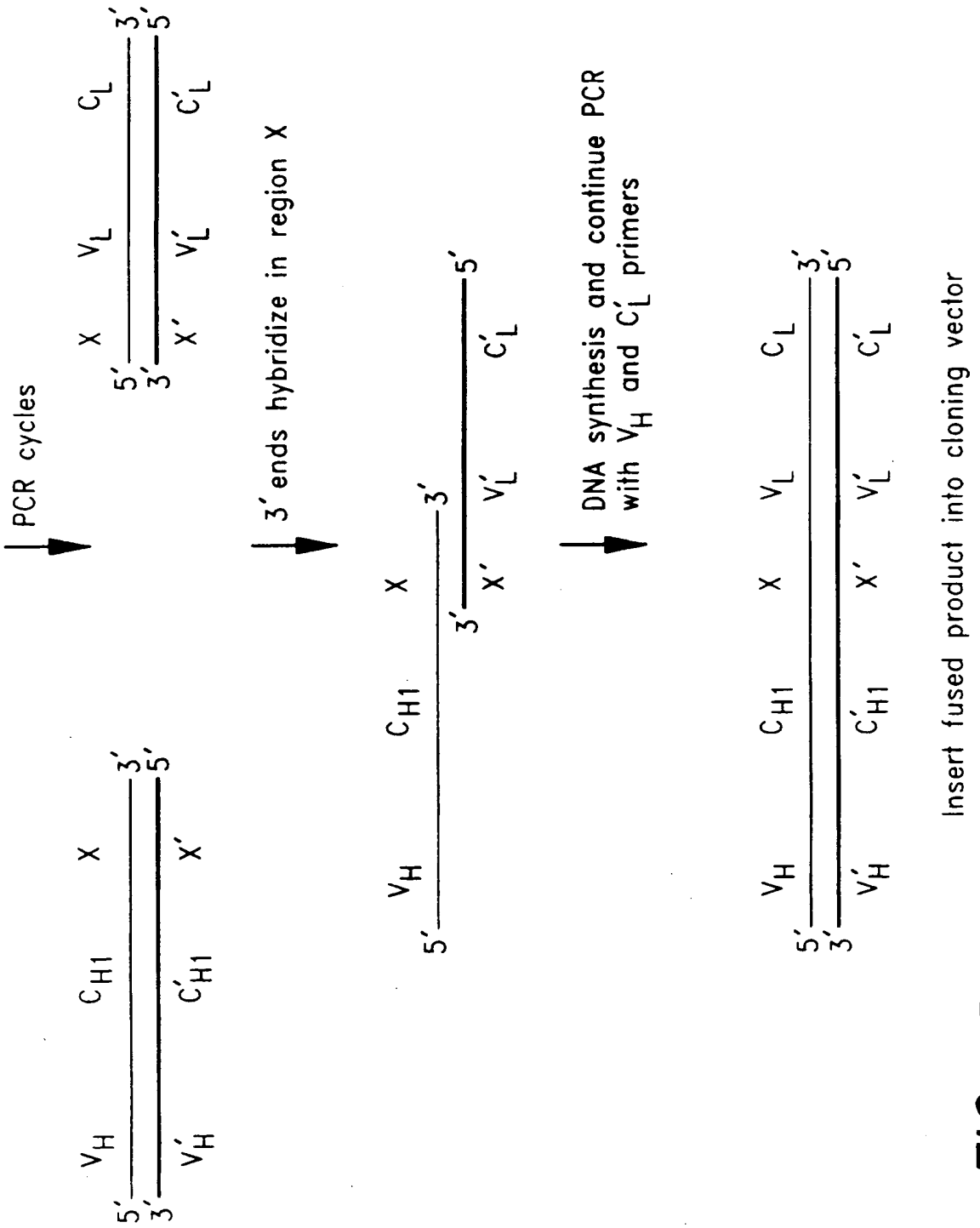


FIG. 4B

heavy chain homology — Spe I stops — rib — pelB leader

M K Y L L P T

5'-GCCTACGG

3' -CTCGGGTTTAGAACATGATCAATTATCGTTCCTCTGTCAGTATTACTTTATGGATAACGGATGCC

heavy chain downstream primer (C_H1')

A A A G L F F I A A Q P A M A Sac I — light chain homology

CAGCCGCTGGATTGTTATTAATCGGCTGCCCAACCTGCCATGGCTGAGCTCGTGATGACCCACTCTCC-3'

GTCGGCGACCTAACAATAA-5'

light chain upstream primer (V_L)

FIGURE 5

V_H EXPRESSION VECTOR:

Not I Ribosome Binding Site

5' GGCCGCAAATTCTATTTCAAGGAGACAGTCATA
CGTTTAAGATAAAGTTCCTCTGTCAGTAT

Pel B Leader

MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA

NcoI XhoI XbaI SpeI
V_H backbone

AlaGlnProAlaMetAlaGlnValLysLeuLeuGlu Thr
GCCCAACCAGCCATGGCCCAGGTGAAACTGCTCGAGATTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACTTTGACGAGCTCTAAAGATCTGA

EcoRI

SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT

FIGURE 6A

V_L EXPRESSION VECTOR:

EcoRI Ribosome Binding Site

```
5'      TGAATTCTAAACTAGTCGCCAAGGAGACAGTCATA
3' TCGAACTTAAGATTTGATCAGCGGTTCCTCTGTCAGTAT
```

Pel B Leader

```
MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeu
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTC
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAG
```

NcoI SacI

```
AlaAlaGlnProAlaMetAlaGluLeu
GCTGCCCAACCAGCCATGGCCGAGCTC
CGACGGGTTGGTCGGTACCGGCTCGAG
```

XbaI
Stop Stop

```
GTCAGTTCTAGAGTTAAGCGGCCG
CAGTCAAGATCTCAATTCGCCGGCAGCT
```

FIGURE 6B

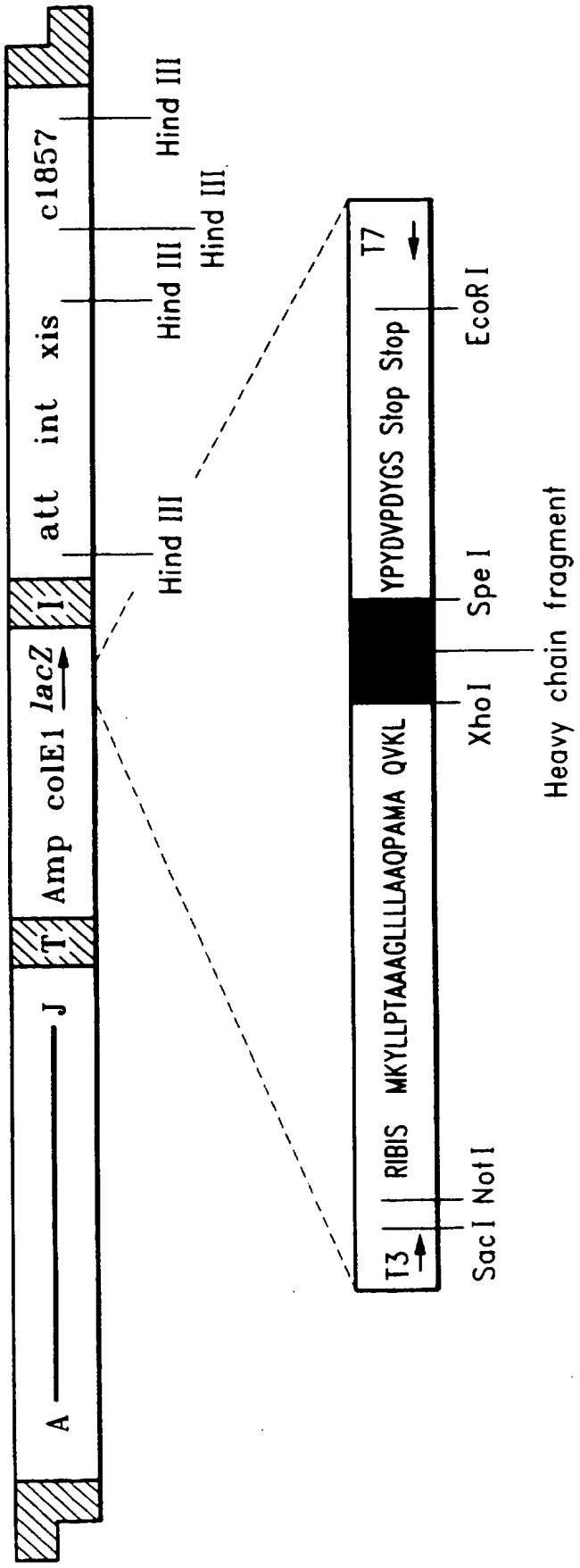


FIG. 7

MODIFIED V_H EXPRESSION VECTOR:

Not I Ribosome Binding Site

```
5' GAGCTGCGGCCGCAAATTCTATTTCAAGGAGACAGTCATA
3'       CGCCGGCGTTTAAGATAAAGTTCCTCTGTCAGTAT
```

Pel B Leader

```
MetLysTyrLeuLeuProThrAlaAlaAlaGlyLeuLeuLeuLeuAla
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCT
TACTTTATGGATAACGGATGCCGTCGGCGACCTAACAATAATGAGCGA
```

NcoI XhoI XbaI SpeI

```
AlaGlnProAlaMetAlaGlnValGlnLeuLeuGlu          Thr
GCCCAACCAGCCATGGCCCAGGTGCAGCTGCTCGAGATTTCTAGACT
CGGGTTGGTCGGTACCGGGTCCACGTCGACGAGCTCTAAAGATCTGA
```

EcoRI

```
SerTyrProTyrAspValProAspTyrGlySerStop
AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAATTCG
TCAATGGGCATGCTGCAAGGCCTGATGCCAAGAATTATCTTAAGCAGCT
```

FIGURE 8A

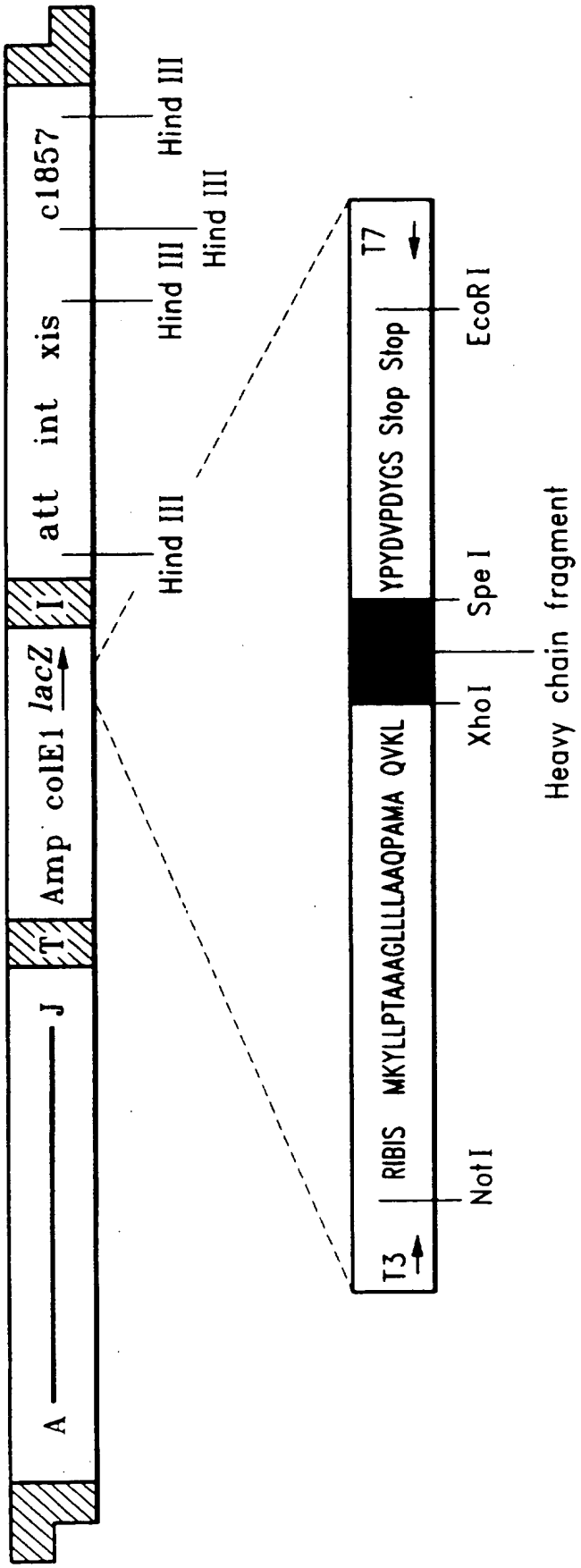


FIG. 8B

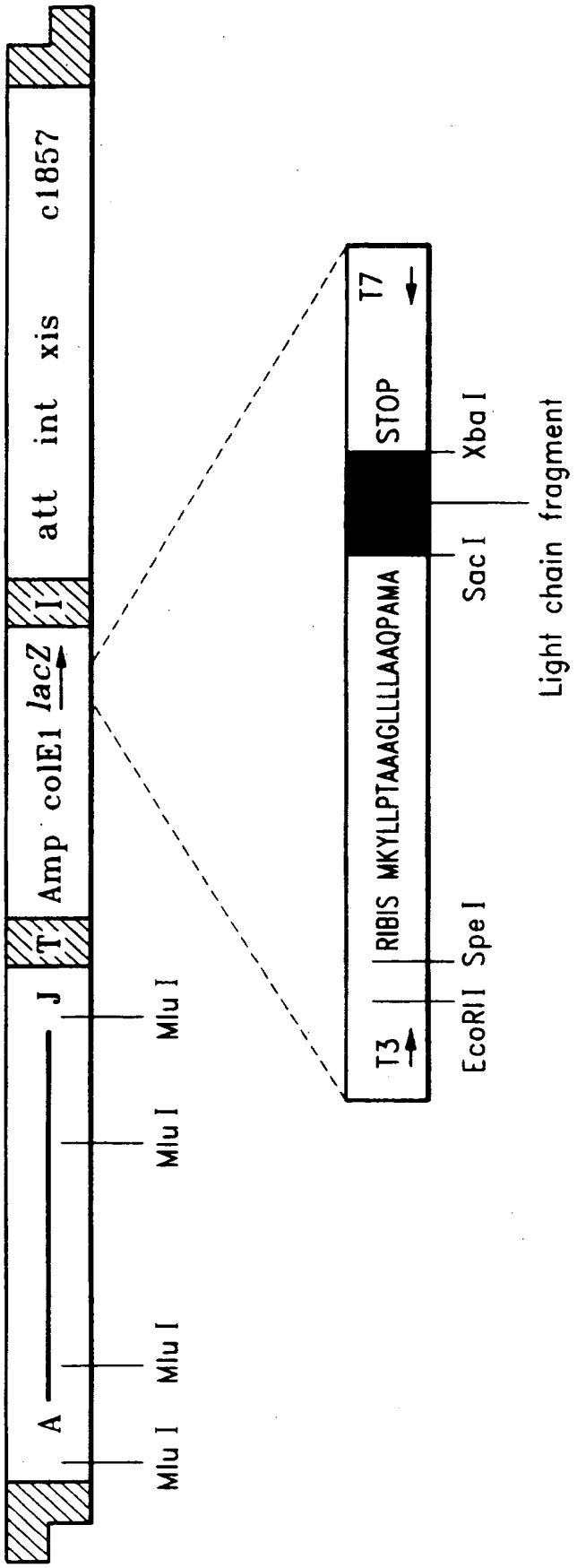


FIG. 9

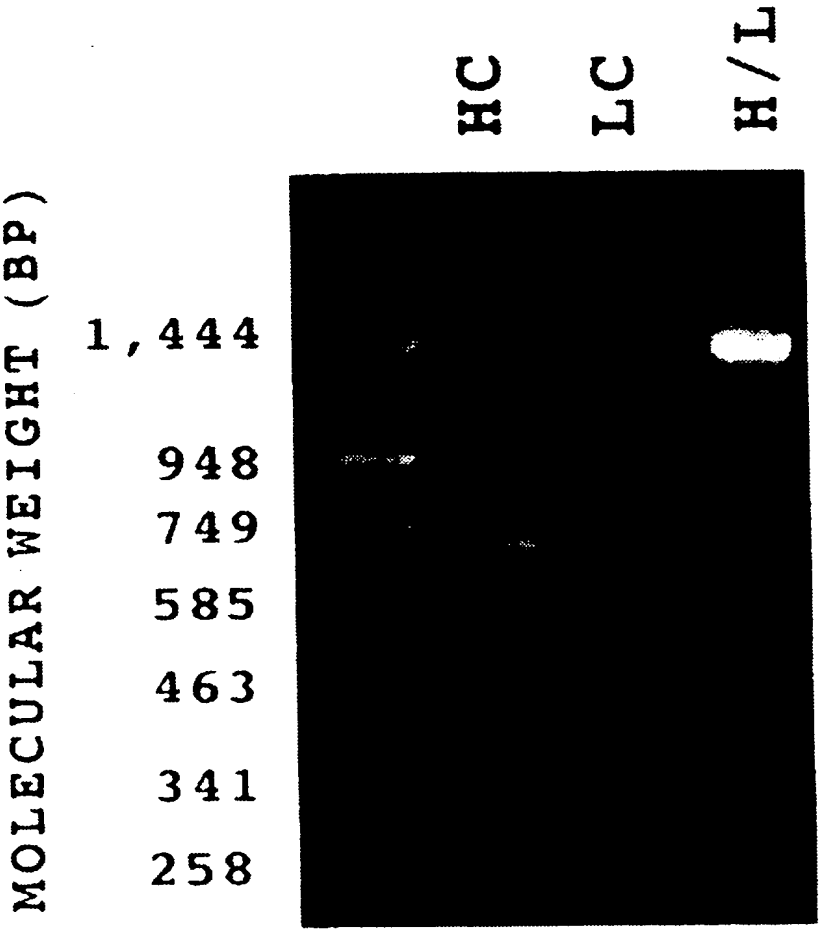


FIG. 10

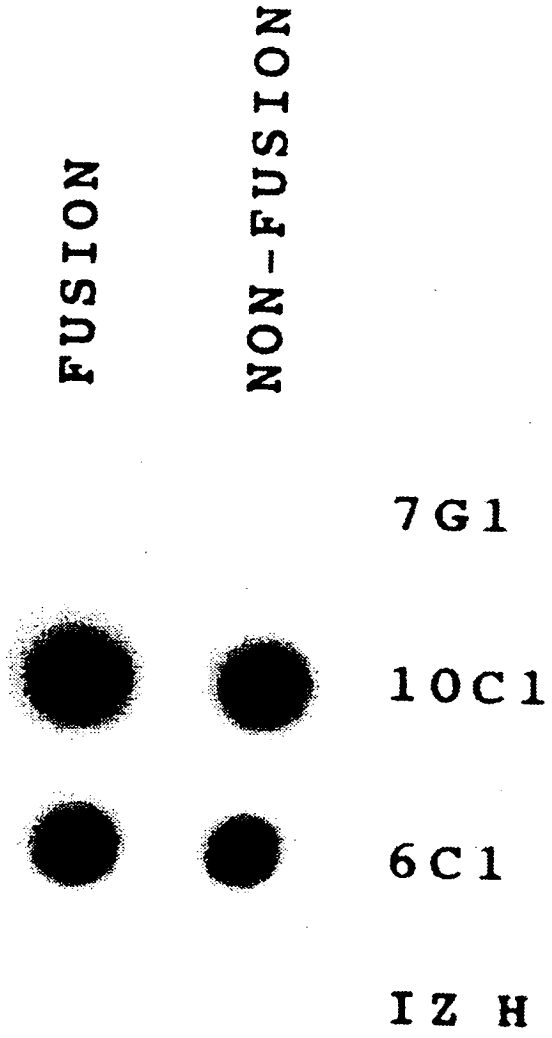


FIG. 11

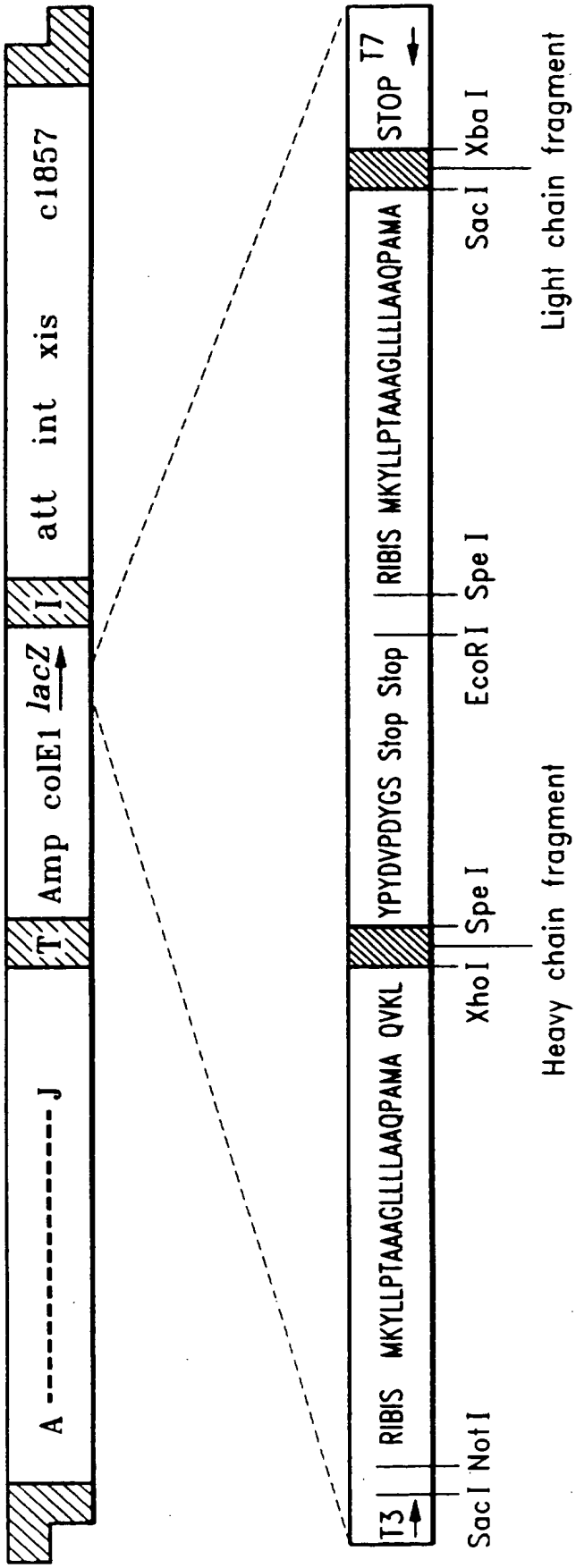

A
J
Amp colE1 lacZ
T
I
att int xis c1857
RIBIS MKYLLPTAAAGLLLLAAQPAMA QVKL
T3
Sac I Not I
Xho I
YPYDVPDYGS Stop Stop
T
Spe I
EcoR I Spe I
RIBIS MKYLLPTAAAGLLLLAAQPAMA
I
Sac I
STOP
T7
Xba I
Heavy chain fragment
Light chain fragment

FIG. 12

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/01475

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/10; C12N15/13; C12Q1/68; //
C12N15/62

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C12Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
P,X	FASEB JOURNAL. vol. 5, no. 6, 19 March 1991, BETHESDA, MD US; A-1717, ABSTR. 7820 R. L. MULLINAX ET AL.: 'ANTIBODY EXPRESSION LIBRARIES IN E. COLI: SIMPLIFIED CONSTRUCTION USING PCR-MEDIATED GENE FUSION' see abstract & 75TH ANNUAL MEETING OF THE FEDERATION OF AMERICAN SOCIETIES FOR EXPERIMENTAL BIOLOGY, ATLANTA, GEORGIA, USA, APRIL 21-25, 1991 ---	1-28
A	WO,A,9 014 430 (SCRIPPS CLINIC AND RESEARCH FOUNDATION) 29 November 1990 ---	

-/--

* Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
09 JUNE 1992	12 JUN 1992
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	THIELE U.H.-C.H.

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	GENE. vol. 77, 1989, AMSTERDAM NL pages 61 - 68; R. M. HORTON ET AL.: 'Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension'	

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9201475 SA 58103

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 09/06/92
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9014430	29-11-90	AU-A-	5673390	18-12-90
		AU-A-	5813890	18-12-90
		CA-A-	2016841	16-11-90
		CA-A-	2016842	16-11-90
		EP-A-	0472638	04-03-92
		EP-A-	0425661	08-05-91
		JP-T-	4500607	06-02-92
		WO-A-	9014424	29-11-90
		AU-A-	5834490	18-12-90
		EP-A-	0478627	08-04-92
		WO-A-	9014443	29-11-90